OLD DOMINION UNIVERSITY RESEARCH FOUNDATION

DEPARTMENT OF PHYSICS AND GEOPHYSICAL SCIENCES
SCHOOL OF SCIENCES AND HEALTH PROFESSIONS
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA

Technical Report PGSTR-AP76-31

# APPLICATION OF SATELLITE DATA AND LARS' DATA PROCESSING TECHNIQUES TO MAPPING VEGETATION OF THE DISMAL SWAMP



*By*

Jeffrey Allan Messmore

Technical Report

*Prepared for the*
National Aeronautics and Space Administration
Office of University Affairs
Washington, DC

*Under*
Grant NGL 47-003-067



May 1976

# APPLICATION OF SATELLITE DATA AND LARS' DATA PROCESSING TECHNIQUES TO MAPPING VEGETATION OF THE DISMAL SWAMP

by

Jeffrey Allan Messmore
B.A. June 1969, Illinois Wesleyan University

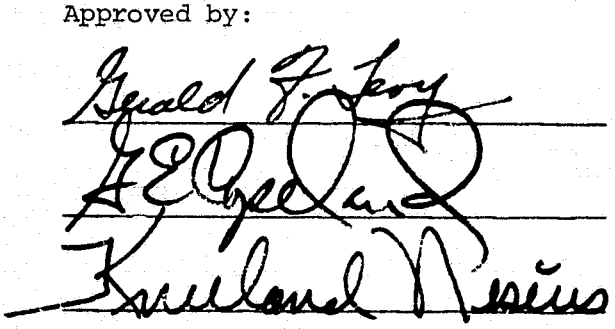
A Thesis Submitted to the Faculty of Old Dominion University
in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

BIOLOGY

OLD DOMINION UNIVERSITY
September, 1975

Approved by:



APPLICATION OF SATELLITE DATA AND LARS' DATA PROCESSING TECHNIQUES TO MAPPING VEGETATION OF THE DISMAL SWAMP.

Jeffrey A. Messmore. Dept. of Biology, Old Dominion University, Norfolk, Va. 23508.

This study concerned the feasibility of using digital satellite imagery and automatic data processing (ADP) techniques as a means of mapping swamp forest vegetation. Multispectral scanner data acquired by the Earth Resources Technology Satellite (ERTS-1; renamed LANDSAT-1) was analyzed using ADP techniques developed by Purdue University's Laboratory for Applications of Remote Sensing (LARS). The site for this investigation was the Dismal Swamp, a 210,000 acre swamp forest located south of Suffolk, Va. on the Virginia-North Carolina border.

Two basic classification strategies were employed in determining the vegetation mapping capability of ERTS-1 data. The initial classification utilized unsupervised techniques which produced a map of the swamp indicating the distribution of thirteen forest spectral classes. These classes were later combined into three informational categories: Atlantic white cedar (Chamaecyparis thyoides), Loblolly pine (Pinus taeda), and deciduous forest. The subsequent classification employed supervised techniques which mapped Atlantic white cedar, Loblolly pine, deciduous forest, water and agriculture within the study site. A classification accuracy of 82.5% was produced by unsupervised techniques compared with 89% accuracy using supervised techniques. Results attained suggest ERTS-1 data and ADP techniques can be successfully applied to mapping some types of swamp vegetation.

ACKNOWLEDGMENTS

Sincere appreciation is extended to the members of my graduate committee: Dr. Gerald F. Levy (Chairman), Dr. Gary E. Copeland, and Dr. Kneeland Nesius for their guidance in this research. In addition, the assistance of Dr. Roger N. Blais, ODU Assistant Professor; Mr. Philip Renfroe, LTV Aerospace Corporation; and Mrs. Virginia Carter, U.S.G.S., is gratefully acknowledged. Also appreciated is the technical assistance provided by personnel at the NASA Langley Research Center and the Laboratory for Applications of Remote Sensing, Purdue University.

This research was in part supported by NASA grant NGL 47-003-067.

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS

# INTRODUCTION

The current trend toward urbanization and continued population growth have increased demands on our forest resources. In order to ensure effective management of this valuable natural asset, a large data base obtained in a timely manner is required. A tool which may prove to be very useful in this area is the Earth Resources Technology Satellite (ERTS-1; now designated LANDSAT-1) launched in July, 1972. Large quantities of digital imagery have been generated, however, the applicability of this data to specific problems remains to be fully investigated. Application of digital satellite data to forest management problems involves the processing of extremely large quantities of data; a task for which computer technology is well adapted.

This study examined the feasibility of applying ERTS-1 multispectral scanner (MSS) data and automatic data processing (ADP) techniques to mapping swamp vegetation. Computer-aided analysis of ERTS-1 data, as applied to forest vegetation mapping, has not been extensively investigated and comparatively little has been published on this phase of satellite data application. Erb (1973) utilized ERTS-1 data to detect, identify, and measure forest and agriculture features of interest. Additionally, automatic classification techniques were applied to a forested area near Cloquet, Minn. by Kirvida and Johnson (1973). Heath and Parker (1973) also used automatic computer processing techniques to map timber stands and range plants in the Houston, Texas area.

The site of this investigation was the Dismal Swamp, a 210,000 acre swamp forest, located within the Virginia cities of Suffolk and Chesapeake, and the North Carolina counties of Currituck, Camden, Perquimans, Gates and Pasquotank (Figure 1). This swamp is one of the largest wilderness

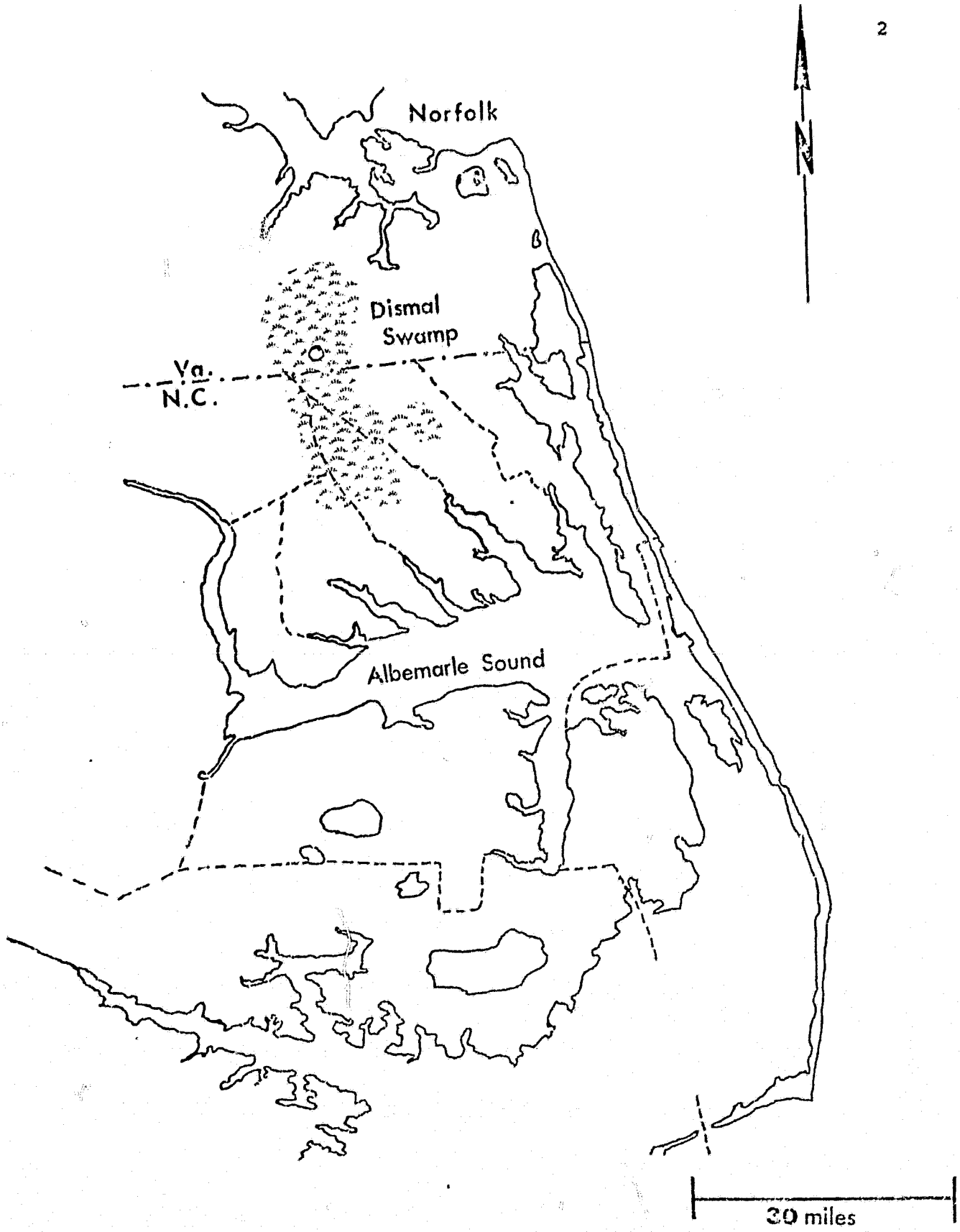



Figure 1. Location of the Dismal Swamp study site.

areas remaining within the Middle Atlantic coastal plain. Most of the swamp is privately owned, except for 49,000 acres which constitute the Dismal Swamp National Wildlife Refuge. Lumbering companies continue to log sizable areas, especially in the North Carolina portion. In addition, forest land is being reclaimed for agricultural purposes in the northeast and southern boundary areas.

Vegetation of the Dismal Swamp was characterized by Meanly (1973), as including five major plant communities: 1) Semihydric or deep water forest, 2) Semihydric or mixed swamp forest, 3) Mesic forest, 4) Atlantic white cedar forest, and 5) Evergreen shrub bog community or pocosin. In 1972, Walker reported the results of extensive quantitative sampling of Dismal Swamp plant communities. She concluded that the swamp's communities exhibit intergradation and species overlap due to the broad tolerances of individual populations. These factors have combined with disturbance to create a complex mosaic of intergrading community types.

Research Objectives

The following objectives were established in order to evaluate the overall utility of ERTS-1 MSS data and automatic data processing techniques as a means of mapping vegetation within the Dismal Swamp.

(1) Evaluate the ability of automatic data processing techniques and ERTS-1 data to accurately differentiate between forested and nonforested areas.

(2) Determine the relative value of the visible (0.5-0.7 μm) and near infrared channels (0.7-1.1 μm) for mapping forest vegetation.

(3) Evaluate the ability of automatic data processing techniques and ERTS-1 data to accurately classify deciduous and coniferous

forest cover.

(4) Determine the extent to which specific tree species or phytocommunities can be accurately mapped using ERTS-1 data and ADP techniques.

Remote Sensing Development

Lindenlaub (1972) defined remote sensing as "the science and art of acquiring information about material objects from measurements made at a distance without coming into contact with the objects". A remote sensing device that is well established as an effective means by which scientists may map plant communities, study successional changes, and identify plant species is aerial photography. Though both black and white and color aerial photography have been used for some time, application of color infrared photography to scientific inquiry is a relatively recent development, growing out of the military's use of infrared film for camouflage detection. The infrared photographic process enables one to visibly record the high reflectivity of vegetation in the near infrared wavelengths. The spectral range of infrared photography is rather limited, having a practical wavelength limit of 1.35 μm. Photographic sensing of longer wavelengths would require cooling the camera and film to the temperature of liquid nitrogen to prevent film fogging (Kodak, 1972). In addition, photographic data is more qualitative than quantitative in nature and therefore is not readily adaptable to automatic data processing. Relatively recent developments in remote sensing hardware have allowed energy recording in the electromagnetic spectrum far beyond the region to which photographic processes are effective. Holter (1970) outlined four principal characteristics of electromagnetic radiation that can be employed in remote discrimination of unknown materials:

(1) spectral variation

(2) polarization variation

(3) spatial variation

(4) temporal variation

A variety of instruments capable of multispectral sensing are available which may be mounted in aircraft or satellites. One type of a sensor that is often utilized in remote sensing systems is the multispectral scanner. This instrument records the radiation reflected and/or emitted from a small area on the earth's surface simultaneously in several discrete regions of the electromagnetic spectrum. Discrimination between target materials is achieved by the differential reflectance or emittance of materials in these regions, or channels, within the energy spectrum. This differential response to radiation in various wavelengths is the foundation upon which remote sensing is built.

Although the electromagnetic spectrum consists of wavelengths which extend from radio waves ($10^{7}$m) to cosmic rays ($10^{-18}$m), the region of most interest to those engaged in remote sensing is from 0.3-15 μm. Hoffer and Johannsen (1969) subdivided this portion of the spectrum into the following regions:

| | |
|---|---|
| (1) 0.3-0.7 μm | optical wavelength region |
| (2) 0.7-3.0 μm | reflective infrared region |
| (3) 3.0-15 μm | emissive or thermal infrared region |

The reflective wavelength range has been further subdivided (LARS, 1970) on the basis of plant spectral response as follows:

| | |
|---|---|
| (1) 0.3-0.72 μm | region where leaf pigment (primarily chlorophyll) predominates |
| (2) 0.72-1.3 μm | very little absorption takes place in this region, and most incident energy is either transmitted or reflected |

(3) 1.3-2.6 μm — water absorption dominates the spectral response of this region

Atmospheric Considerations

Local weather conditions, as well as variation in illumination, can affect the nature and quality of remotely sensed data. Atmospheric phenomena cause attenuation losses that are the result of scattering, absorption, and reflection of radiation. As shown in Figure 2, ozone found in the upper atmosphere strongly absorbs the ultraviolet portion of the spectrum while electromagnetic energy at the infrared end is gradually reduced by water vapor and $CO_2$ absorption. In order to avoid the strongly attenuated wavelengths, sensor data is acquired through atmospheric windows, or regions within the energy spectrum where a large percentage of the total emitted or reflected radiation is transmitted.

Plant Reflectance

Radiant energy is reflected and transmitted by a plant leaf in a manner that is characteristic of pigmented cells containing water solutions. Other factors which affect the spectral quality and intensity of plant reflectance and emittance are leaf geometry, morphology, physiology, chemistry, soil site, and climate (Gates, 1970). The reflectance spectrum of a typical mature green leaf is shown in Figure 3. It shows high leaf absorption in the ultraviolet and blue portions of the electromagnetic spectrum, comparatively low absorption in the green and high absorption again in the red wavelengths. In the near infrared (0.7-1.3 μm) there is very low absorption, and thus high reflectance and transmittance followed by very high absorption in the infrared at 1.45 μm and 1.95 μm.

Energy absorption within leaf pigment molecules takes place as a result of electron transitions which require high energy photons. Pigment molecules do not absorb the near infrared wavelengths which are of lower

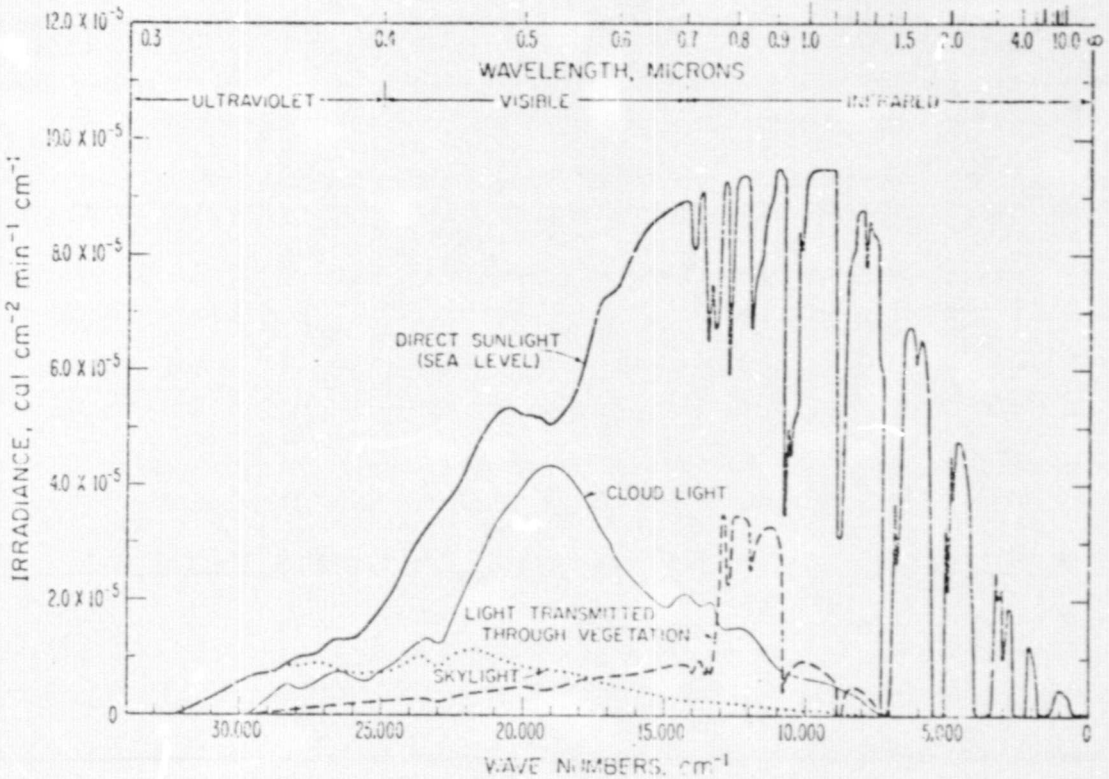



Figure 2. Energy distribution of direct sunlight, cloud light, light transmitted through vegetation, and skylight plotted as a function of wave number (Gates, 1965).

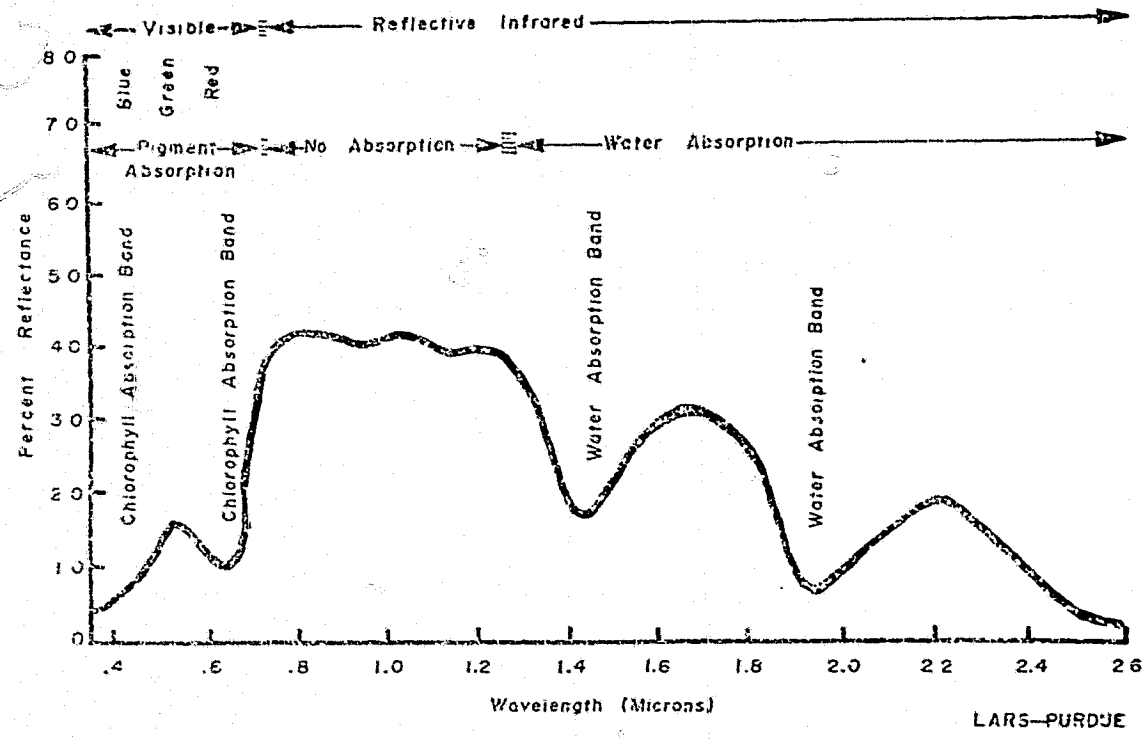



Figure 3. Reflectance spectrum of a mature green leaf (LARS, 1970).

energy content. At longer infrared wavelengths, water and other molecules within the leaf dominate the leaf spectral response as they undergo low energy vibrational and rotational quantum transitions (Gates, 1967). Energy absorbed by the leaf in the visible wavelengths is utilized within the leaf for photosynthesis and the production of plant nutrients. Energy absorbed between 0.7-1.3 μm, however, is converted into heat within the leaf. Fortunately, a leaf reflects and transmits the near infrared well and absorbs relatively little. In this way a sunlit leaf is able to remain substantially cooler than it would if it maintained high absorption in the near infrared; a high energy region where over 50% of incident sunlight energy is concentrated (Gates, et al. 1965).

According to Myers (1970), leaf reflectance in the near infrared to 1.3 μm is due primarily to the physiological structure of the plant leaf. Gates et al. (1965) felt that near infrared leaf reflectance was probably a function of cell shape and size as well as the amount of intercellular space. Gates' hypothesis is largely supported by Woolley (1971) who reported that diffuse cellulose reflectance is primarily responsible for the infrared reflectance character of a dried leaf, while the infrared reflectance curve of a fresh leaf is dependent on a combination of diffuse cellulose reflectance and $H_2O$ absorption bands. While working with salinity stressed cotton, Gausman et al. (1969) found that changes in internal leaf structure were the primary causes for differential reflection response in the 0.75-1.3 μm range. They associated increased reflectance with thicker leaves, having more pronounced palisade development, loosely arranged spongy mesophyll, and consequently more cellular interfaces where reflectance might occur.

Leaf orientation to incident sunlight can significantly alter the

intensity of reflected spectral radiation in some regions of the electromagnetic spectrum. Cotton leaf stacking experiments by Myers et al. (1966) have shown that in the visible wavelengths, using any combination of stacked leaves, there is no change in reflectance. This evidence indicates that the topmost exposed leaves are the only ones responsible for reflectance intensity in the visible wavelengths. However, in the near infrared, reflectance of two leaf layers increased approximately 17% over that of one layer. As the number of leaf layers increased up to six layers, reflectance increased. Myers felt the probable explanation for this phenomena is that light transmitted through the top leaves is partly reflected from and partly transmitted through the lower leaves. Infrared radiation reflected upward from the lower leaves augments the reflected light from the surface of the leaves. One must be cautious, however, when applying laboratory derived data to natural scenes. This was demonstrated by Stoner and Baumgardner (1972) when they reported that for agricultural crops, canopy reflectance is much less than for a single leaf because other scene parameters such as illumination angle, leaf orientation, shadows, and non-foliage background (primarily soil) cause reflectance attenuation. Knipling (1969) estimated that visible reflectance from a crop canopy may be 3-5% while that for a single leaf is 10%. He also noted a 30% reduction in the near infrared reflectance of a complete canopy as compared to the reflectance of a single leaf. This observed reduction in the visible and near infrared reflectance can be accounted for when the previously mentioned attenuating phenomena are considered.

## METHODS and MATERIALS

In July of 1972 the National Aeronautics and Space Administration (NASA) launched ERTS-1 as part of a program to demonstrate the applicability of remote sensing to the management of earth's resources. The satellite was launched in a circular, sun synchronous, near-polar orbit at an altitude of 494 nautical miles. Every 18 days at the same local time the satellite's ground trace repeats its coverage. Optical energy is sensed by a four channel multispectral scanner which uses an oscillating mirror to continuously scan perpendicular to the spacecraft velocity in swaths of 100 nautical miles (NASA, 1972). A single resolution element (pixel) within that swath is rectangular and measures approximately 60 x 80 m (Todd et al. 1973). Electromagnetic energy is sensed simultaneously by an array of detectors in four wavelength regions (Table 1).

On August 30, 1973, ERTS-1 imaged an area 100 x 100 nautical miles in the middle Atlantic Coast region of the United States. Contained within this image was the Dismal Swamp. Data collected over the swamp was analyzed with automatic data processing techniques developed by Purdue University's Laboratory for Applications of Remote Sensing (LARS). Analysis involved the use of a remote computer terminal located at the NASA Langley Research Center, Hampton, Va. This terminal was tied into an IBM system/360 model 67 computer at Purdue University. Data processing required the utilization of six LARS computer programs, normally performed in the following sequence: 1) Pictureprint 2) Cluster 3) Statistics 4) Feature selection 5) Classifypoints and 6) Printresults (Figure 4).

### Classification Strategies

Both supervised and unsupervised classification strategies were employed in mapping vegetation within the swamp. The basic difference between these

Table 1. Spectral range of ERTS-1 bands and corresponding LARS channels.

| ERTS-1 Band | LARS Channel | Spectral Range (μm) |
|---|---|---|
| 4 | 1 | 0.5 - 0.6 |
| 5 | 2 | 0.6 - 0.7 |
| 6 | 3 | 0.7 - 0.8 |
| 7 | 4 | 0.8 - 1.1 |

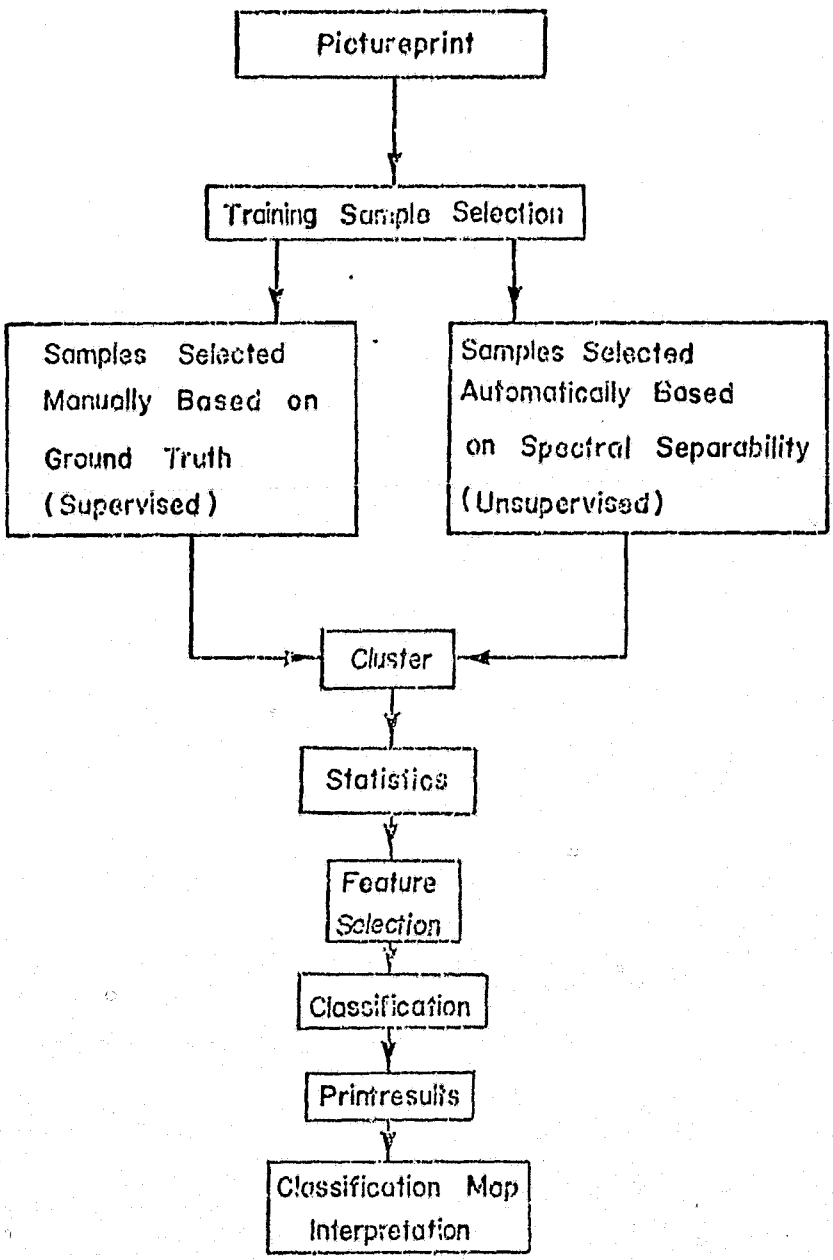



Figure 4. LARS data processing sequence (revised from Coggeshall and Hoffer, 1973).

two methods lies in computer training (See Figure 4). When using supervised classification techniques, ground truth is available and the computer is trained to recognize known materials by their spectral response. In contrast, unsupervised classification techniques are employed when ground truth is not available or is insufficient to develop a supervised classification. When using this method, the computer is trained wholly on the data's spectral composition.

The supervised classification of data acquired over the swamp involved first identifying areas of interest where ground truth information was available at the time of data acquisition, then defining these areas to the computer. After training samples were selected, statistical parameters which characterized the spectral response of known ground features were computed. These were subsequently utilized by the pattern recognition algorithm to classify each data point within the swamp into one of the designated material classes.

Unsupervised classification began by defining as training fields those areas within the study site that contained representative ground features. This information was divided into a number of clusters, i.e. spectral classes, based on the distribution of spectral information within the training field data. The reflectivity of each cluster was then defined statistically. Data points representing unknown ground features within the swamp were subsequently classified into one of the statistically defined spectral classes by a pattern recognition algorithm. Final analysis centered upon determination of the relationship between the mapped spectral classes and known surface data, i.e. ground truth comparison.

Pictureprint

The first program run in the analysis sequence was the pictureprint

function. This program produces a gray scale printout resembling a low resolution photograph (Figure 5). Each pixel within the printout is located by a specific column and line number. The pictoral character of the printouts is attained by assigning alphanumeric symbols to radiance levels within each channel. Highly reflective areas were assigned symbols which covered a low percentage of the printout, e.g. ---- and absorbent areas dense appearing symbols, such as MMM. Because the data are not evenly distributed between the highest and lowest radiance values, symbols cannot be assigned so that each represents an equal range of radiance. Instead, to achieve maximum contrast within the pictureprint, the data are histogrammed and symbols are assigned bins of radiance values so that each will occur with approximately the same frequency (Figure 6).

The gray scale printouts were often very useful for identifying areas of known composition within the data. These areas could then be designated as ground truth sites and used to train the computer to recognize similar materials. For example, training fields in agricultural investigations may correspond to specific crop species, such as oats or corn; while in forestry studies, coniferous and deciduous forest, or pure stands of a tree species may be designated. Usually only some of the ground truth areas were used to train the computer; remaining areas were employed later for testing the accuracy of the completed computer classification. Training fields were located through the designation of column and line numbers corresponding to natural boundary lines. If these boundaries were not rectangular in shape, a number of small rectangular fields were used to define the desired training area.

Cluster

The next program employed in this analysis was the cluster function.

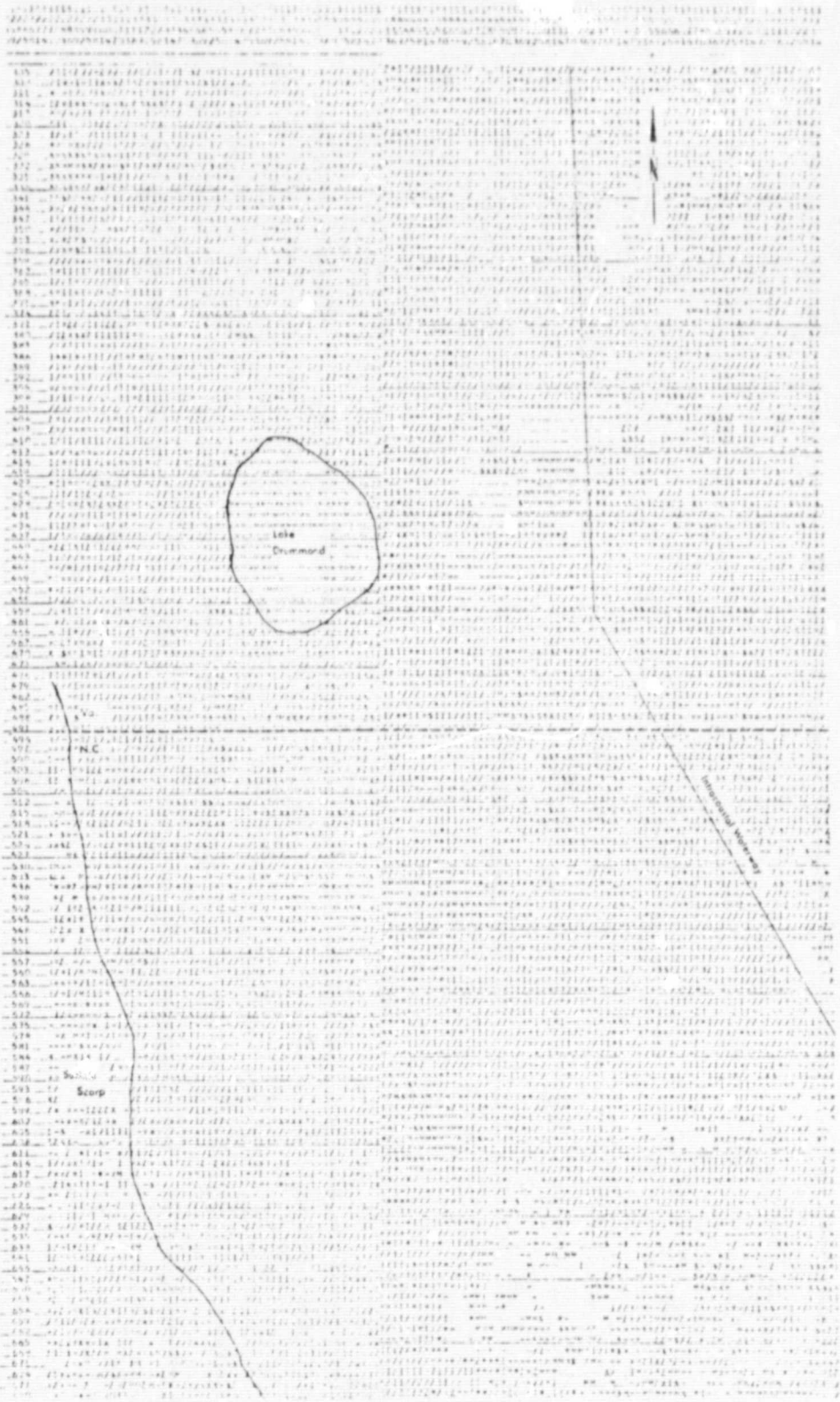



Figure 5. Pictureprint of a portion of the Dismal Swamp (channel 4 data).

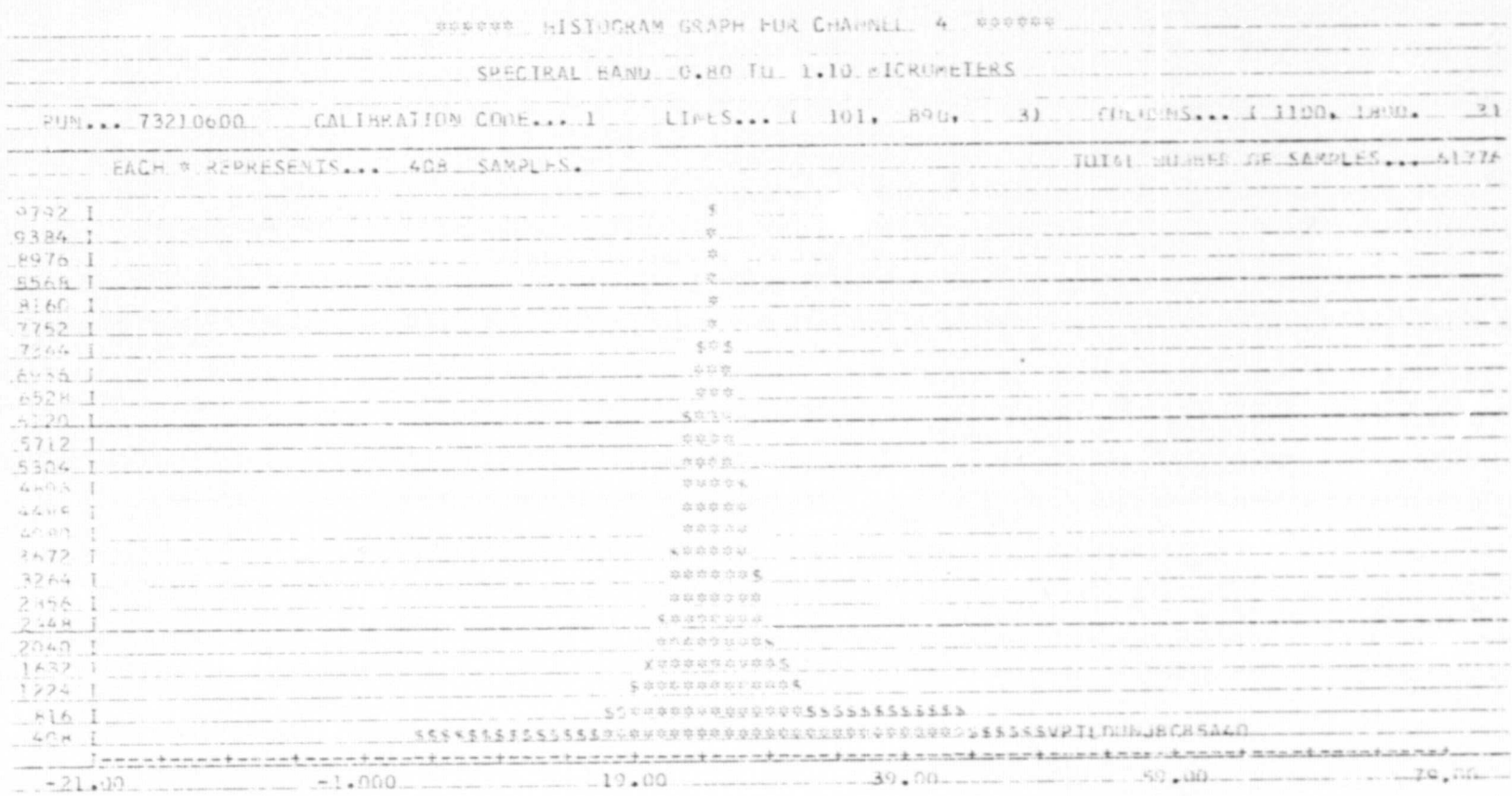



Figure 6. Histogram of spectral data collected from the swamp.

This program was necessary because the pattern recognition algorithm assumes that each class can be characterized by a multidimensional probability function. The researcher must determine if this assumption is a valid one for each of the chosen training classes. Clustering the training samples for each class, e.g. Loblolly pine, Atlantic white cedar, etc. determines whether or not the multispectral data is Gaussian (unimodal) in nature and provides a means for dividing the training classes into approximate Gaussian sub-classes if the original data is non-Gaussian (Lindenlaub, 1973). Cluster function options allow the researcher to choose any combination of available spectral bands, as well as the number of cluster groups to be separated within the data. Lindenlaub (1973) suggests, however, that the analyst designate a number of spectral groups which is twice the expected number of separable clusters. If one specifies a larger number of spectral clusters than is actually present, the computer will reduce this number sequentially until the maximum number actually present in the data has been found. During the clustering operation, training class data were examined, and the entire set of data were statistically divided into a number of groups; each containing data points having similar spectral characteristics (Hoffer et al. 1972). Preprogrammed alphanumeric symbols were used to map the cluster groups within the training data. In addition to the printing of cluster maps, tables containing the means and variances of each class, as well as the pairwise separability values between all class pairs were listed. These data were useful in determining the number of spectral clusters that could reasonably be separated within the training data. Lindenlaub (1973) suggested using the pairwise quotient value of 0.8 as a rule of thumb breakpoint for separability. He

emphasized, however, that this breakpoint, although pretty well established through extensive experience with agricultural data, is problem and data dependent. If the separability data indicated that some clusters were not sufficiently separable, these spectrally similar groups were combined. It was often found that a training field composed of an apparently homogenous ground cover could be better represented by three or four classes in order to more accurately approximate a Gaussian data distribution.

Statistics

The pattern recognition algorithm classifies data on the assumption that spectral information derived from the various training classes can be characterized by a Gaussian probability density function. Each density function can therefore be defined by its mean vector and covariance matrix. All four channels were employed in estimating each training classes' mean vector and covariance matrix.

In addition to computing mean vectors and covariance matrices, the statistics function produced training class histograms in all ERTS-1 channels. These histograms served as a partial check as to whether the training class multispectral data was Gaussian in nature. Lindenlaub (1973) reported that the histograms are only a partial check on the data distribution because they depict only the marginal density functions and do not necessarily represent the multidimensional density function as a whole. If a class histogram shows a multimodal pattern, the data is probably non-Gaussian and indicates that further refinement of the training class is necessary.

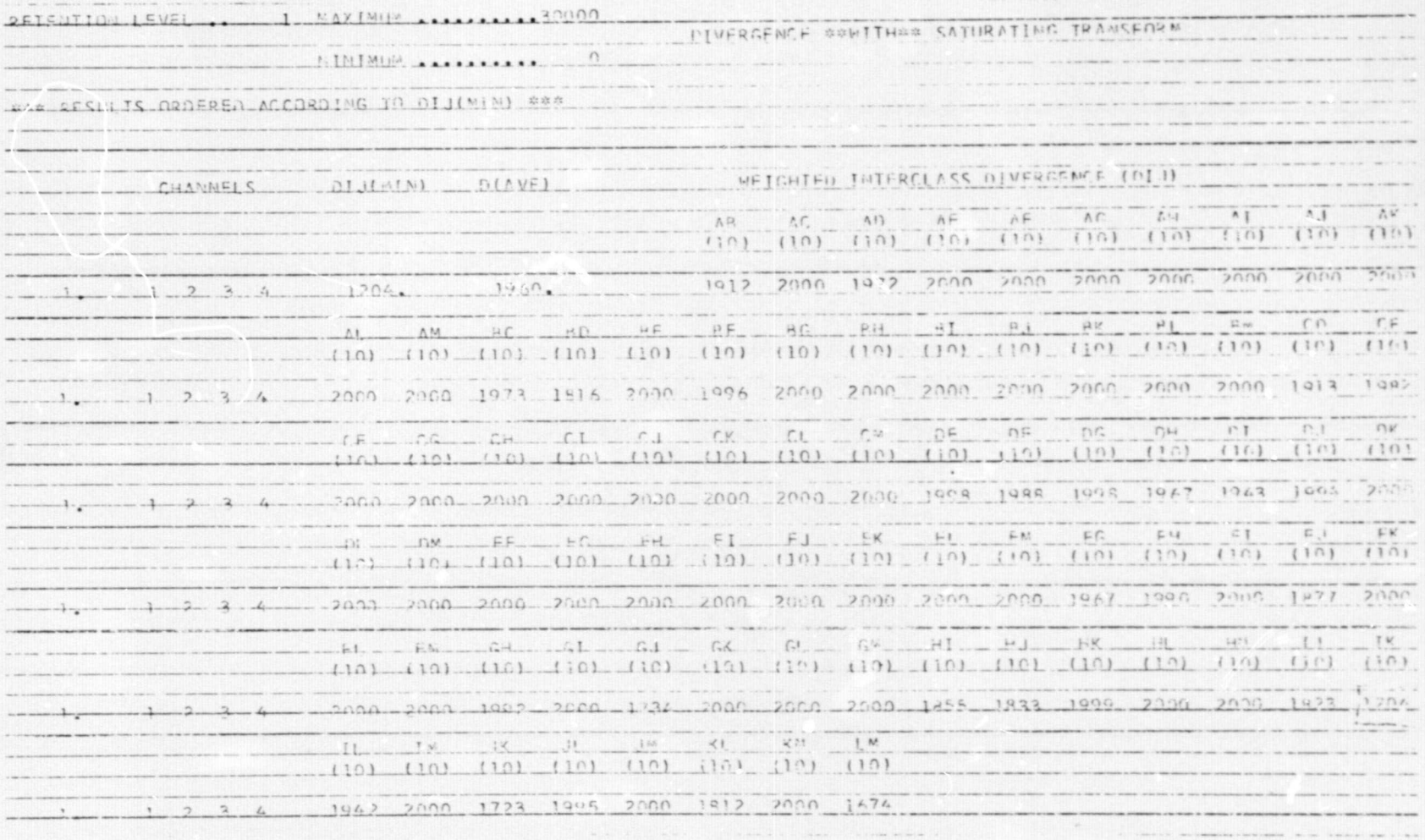
Feature Selection

This processing function was used to determine which features, or channels, within the scanner data were best to use in the final classification.

This was not an acute problem with four channel ERTS-1 data because of the small number of channels available. However, with twelve channel aircraft data, studies have shown that as few as four or five channels may be utilized without seriously compromising classification accuracy (Lindenlaub, 1973). Use of additional features in the classification step substantially increases the amount of computer time necessary. This program calculated the statistical distance in N-dimensional space (N = number of channels) between the training classes that had been provided it. The requested channel combinations were then ranked in terms of the average or minimum pairwise distance between all class pairs. Flexibility is added to the program through the capability of differentially weighing designated class pairs in computing the average or minimum distance between classes. This could be used, for example, if the analyst had found four spectral classes within forest, though the classification objective was to map forest as a single entity. Since a classification mistake between the four forest classes would be immaterial to classification accuracy, the distance between these classes would be given a lower weight. As can be observed in the sample feature selection output, (Figure 7) interclass divergence (a measure of class dissimilarity) saturates at a value of 2000. However, Lindenlaub (1973) reported that statistical distances on the order of 1700 or larger will generally yield satisfactory classification accuracies. If it was found that the statistical distance between significant materials was low, it was sometimes necessary to repeat one or more of the previous analyses in an attempt to increase the critical divergence values.

Classification

Study site classification, the last step in the analysis sequence,



DIVERGENCE **WITH** SATURATING TRANSFORM

*** RESULTS ORDERED ACCORDING TO DIJ(MIN) ***

| CHANNELS | DIJ(MIN) | D(AVE) | WEIGHTED INTERCLASS DIVERGENCE (DIJ) |
|---|---|---|---|
| | | | AB (10) AC (10) AD (10) AE (10) AF (10) AG (10) AH (10) AI (10) AJ (10) AK (10) |
| 1, 1 2 3 4 | 1204. | 1960. | 1912 2000 1932 2000 2000 2000 2000 2000 2000 2000 |
| | | | AL (10) AM (10) AC (10) AD (10) AE (10) AF (10) AG (10) AH (10) AI (10) AJ (10) AK (10) AL (10) AM (10) AN (10) AO (10) AP (10) |
| 1, 1 2 3 4 | 2000 | 2000 1973 1816 2000 1996 2000 2000 2000 2000 2000 2000 2000 1913 1982 | |
| | | | CE (10) CG (10) CH (10) CI (10) CJ (10) CK (10) CL (10) CM (10) DE (10) DF (10) DG (10) DH (10) DI (10) DJ (10) DK (10) |
| 1, 1 2 3 4 | 3000 | 2000 2000 2000 2030 2000 2000 2000 1998 1988 1996 1967 1963 1993 2000 | |
| | | | DL (10) DM (10) EE (10) EG (10) EH (10) EI (10) EJ (10) EK (10) EL (10) EM (10) EN (10) EO (10) EP (10) EI (10) EJ (10) EK (10) |
| 1, 1 2 3 4 | 2000 | 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 1967 1990 2000 1977 2000 | |
| | | | EL (10) EM (10) EN (10) EO (10) EP (10) EQ (10) ER (10) ES (10) ET (10) EU (10) EV (10) EW (10) EX (10) EY (10) EZ (10) FK (10) |
| 1, 1 2 3 4 | 2000 | 2000 1982 2000 1934 2000 2000 2000 1955 1933 1999 2000 2000 1923 1200 | |
| | | | FL (10) FM (10) FN (10) FO (10) FP (10) FQ (10) FR (10) FS (10) |
| 1, 1 2 3 4 | 1942 | 2000 1723 1995 2000 1812 2000 1674 | |

Figure 7. A portion of the feature selection output obtained during unsupervised classification.

was the culmination of all previous analyses. Inputs to the program included a statistics deck, the selected combination of channels, and coordinates of the area to be classified. The pattern recognition algorithm was used to individually classify each pixel into one of the statistically defined training classes on a maximum likelihood basis. Test fields were designated to estimate classifier performance. These fields were examined and each data point was individually classified into its most likely training class. Results of this process were employed to determine classification accuracy.

Output of the classification function, combined with a printresults program, produced a classification map with alphanumeric symbols representing designated material classes. The computer also calculated how accurately it classified training data by comparing the classification of each point in the training fields with the initial ground truth designation. A high level of agreement indicated there was little confusion within the training statistics for the various designated materials, and the classes were being accurately separated. When agreement in test and training field performance was low, redefinition of the classes was necessary.

Thresholding was utilized during the generation of classification maps. If this option had not been applied, the classification algorithm would have included every data point into the class it most nearly resembled, even though resemblance was remote. Thresholding allows the researcher to arbitrarily screen out those picture elements not demonstrating a high degree of correlation with user-designated spectral classes. Thresholded points appear as blank spaces on the final classification map.

## RESULTS and DISCUSSION

### Unsupervised Classification

The initial analysis step produced gray scale printouts of the study area in all four ERTS-1 channels. These printouts were compared with black and white, color and color IR aerial imagery of the swamp. Pictureprints of the data in channels 1 and 2, 0.5-0.6 μm and 0.6-0.7 μm respectively, produced the best delineation of swamp boundaries with the Suffolk scarp on the swamp's western edge being clearly defined. In addition, U.S. route 460 and the Norfolk and Western railway were discernable cutting east to west across the northern region of the swamp. Pictureprints produced of channels 3 and 4, 0.7-0.8 μm and 0.8-1.1 μm respectively, were especially useful in the delineation of water, due to its high absorbancy in these wavelengths. Lake Drummond, located in the swamp's center was clearly differentiated from the surrounding forest, however, none of the numerous drainage ditches located in the swamp could be identified. This is believed to be due to tree overhang and the characteristic low water levels of the swamp in August.

In order to determine the distribution of spectral information collected from the swamp, the clustering algorithm was run on eleven training fields (15,894 data points). Cluster results maps were then printed which contained alphanumeric symbols indicating the geographic location of spectrally similar materials (Figure 8). Clustering performed with all four wavelength channels was not found to be optimal for obtaining spectral separability within the forested training area. Comparison of the cluster means and standard deviation values in all four channels showed the cluster radiance means in channels 1 and 2 to be nearly identical, thereby causing a large amount of data overlap in two of the four clustering


Figure 8. Forest training field cluster map. Symbols indicate the location of spectrally similar materials.

space dimensions. This resulted in the less than optimal separability.

Elimination of channels 1 and 2 in the clustering process increased the number of separable clusters. A comparison of the separability quotient values for a four channel, 12 cluster run and one using the two IR channels alone, yielded fifteen separability quotient values of less than 0.75 for the former and only one value of less than 0.75 for the latter. A value of 0.75 is usually used as the "breakthrough" point. Results obtained from other cluster runs are illustrated in Figure 9.

Further clustering analysis of the two IR channels suggested the presence of thirteen spectral groups. These were submitted to the statistics processor which calculated the mean vector and covariance matrix for each of the proposed forest classes. The mean reflectance vector for each forest class is presented in Table 2. In order to examine the distribution of data within each class, this processor was also requested to graph spectral histograms in all four channels. Examples of these class histograms are shown in Figure 10. In every case an approximate Gaussian distribution was exhibited. A coincident spectral plot of all four ERTS-1 channels was also produced which illustrates the relative amplitude of each classes' spectral response (Figure 11). Each bar on this graph is proportional to the mean class spectral response, + and - one standard deviation. As expected, channels 1 and 2 were incapable of differentiating the forest classes due to radiance value overlap. Channels 3 and 4, however, contained a wider range of mean spectral values and, therefore, the ability to distinguish the thirteen classes.

Feature selection was subsequently utilized to determine the combination of available channels that would yield the most accurate classification results with a minimum amount of computer time. This function

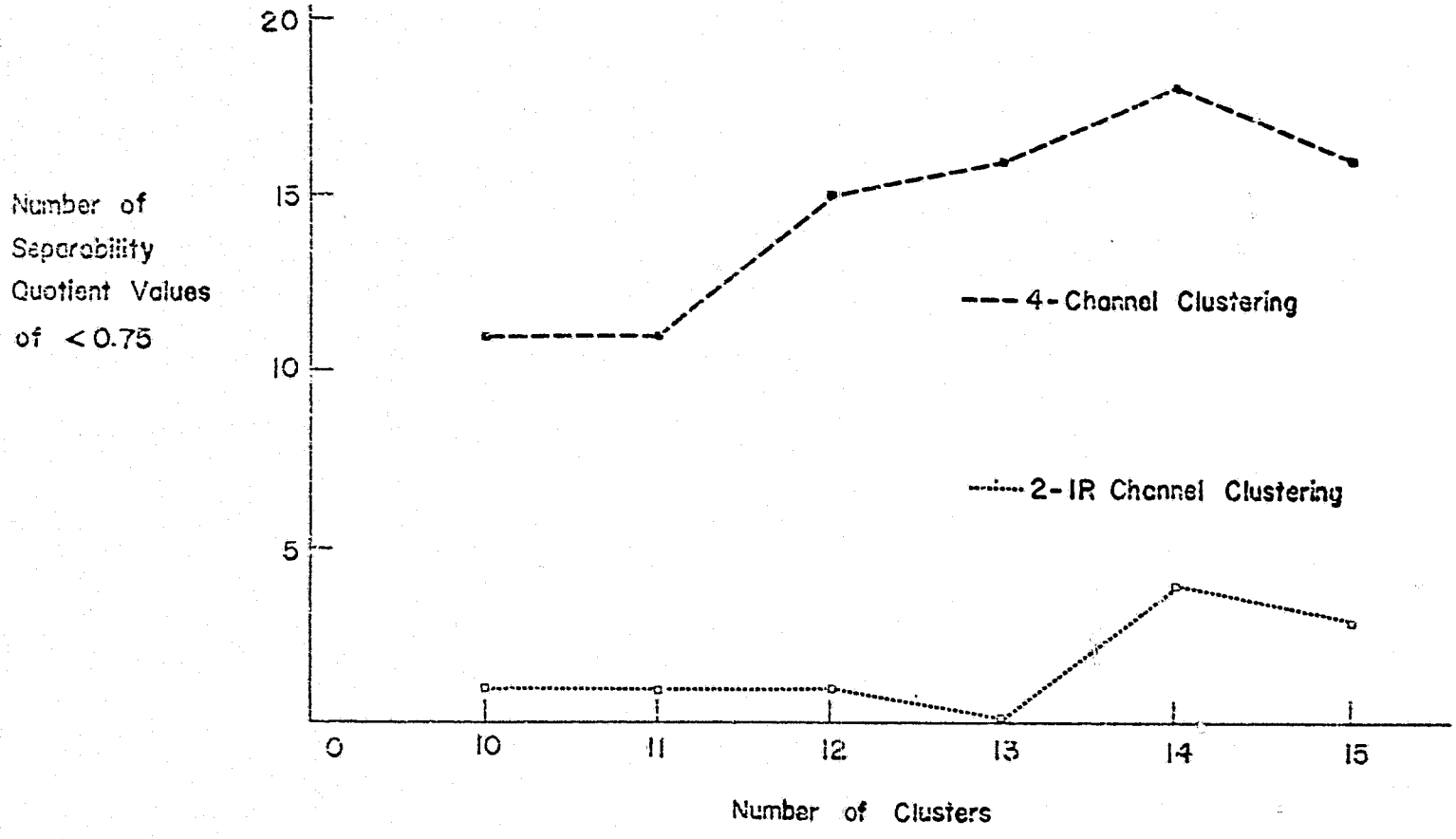



Figure 9. Results of clustering performed with all four channels and with the infrared channels only.

Table 2. Mean reflectance values for thirteen forest spectral classes.

| SPECTRAL CLASS | MEAN REFLECTANCE VALUES | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 26.08 | 15.44 | 45.42 | 26.29 |
| 2 | 25.72 | 15.02 | 42.44 | 24.93 |
| 3 | 25.09 | 14.70 | 40.96 | 24.60 |
| 4 | 26.38 | 15.92 | 41.11 | 22.25 |
| 5 | 25.01 | 14.80 | 40.01 | 23.59 |
| 6 | 24.89 | 14.58 | 38.50 | 24.18 |
| 7 | 24.77 | 14.80 | 38.50 | 22.98 |
| 8 | 25.34 | 15.34 | 38.65 | 21.65 |
| 9 | 25.45 | 15.11 | 37.50 | 20.24 |
| 10 | 24.73 | 14.63 | 36.60 | 22.33 |
| 11 | 24.81 | 14.84 | 35.48 | 20.57 |
| 12 | 25.08 | 15.20 | 33.22 | 18.88 |
| 13 | 24.96 | 15.29 | 29.38 | 16.07 |

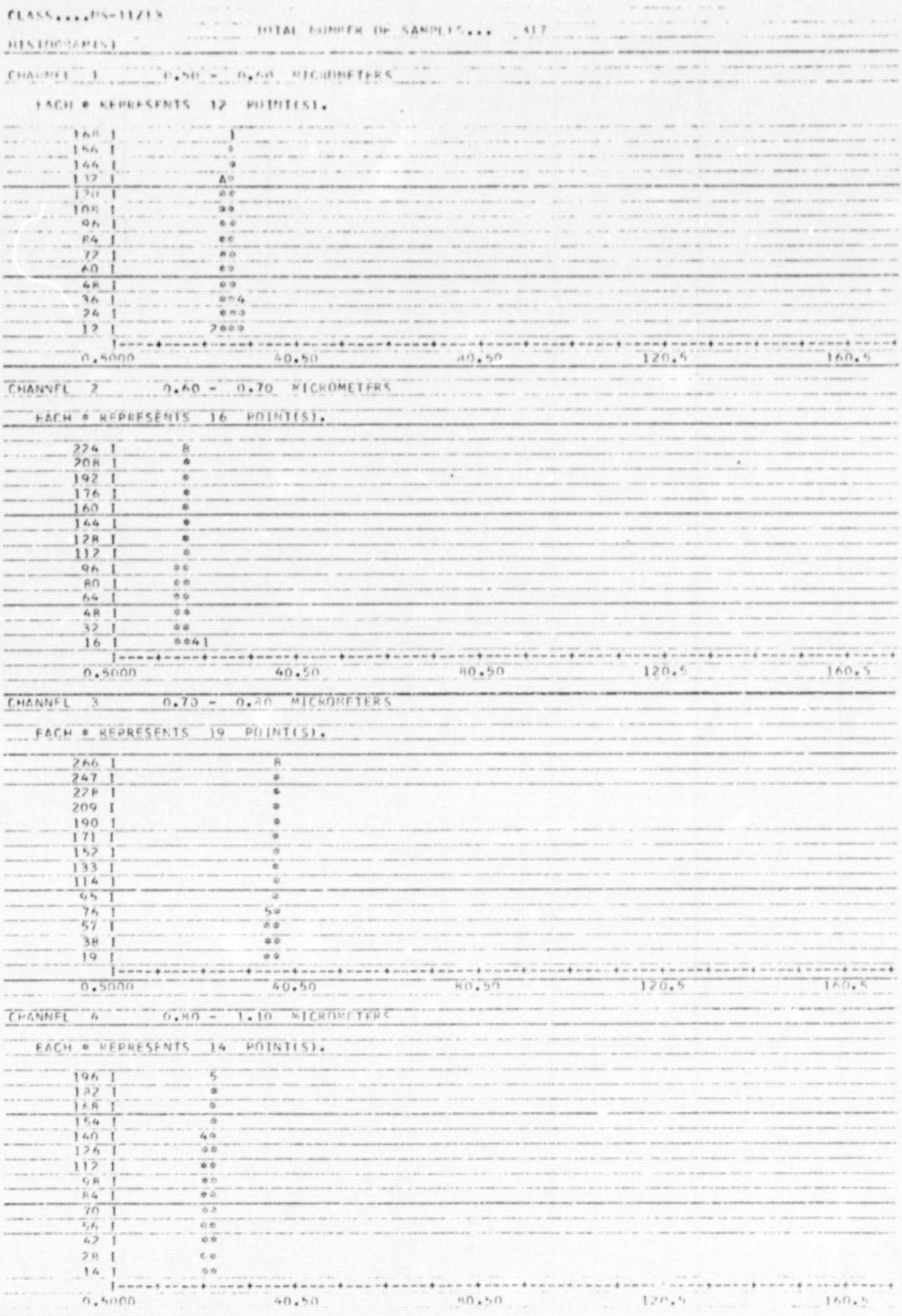



Figure 10. Histograms of the spectral data in forest class 11.

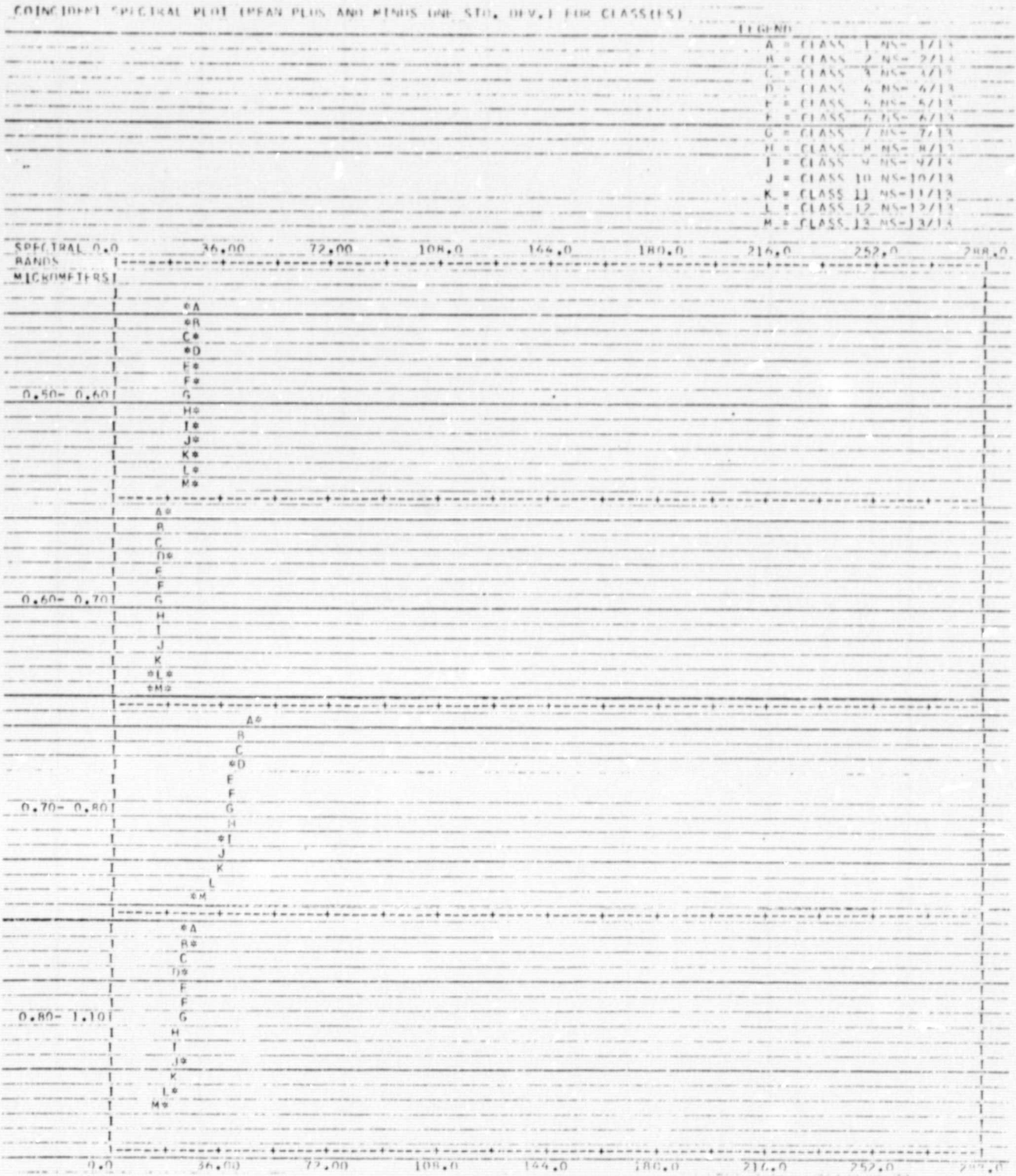



Figure 11. Coincident spectral plot of thirteen forest classes.

calculates the statistic transformed divergence, which is a "measure of the dissimilarity of two distributions" and "provides an indirect measure of the ability of the classifier to discriminate successfully between them" (Swain 1973). The strategy utilized was to weight all classes equally and maximize the pairwise transformed divergences. Table 3 presents a summary of the feature selection output. The combination of all four channels performed best, producing the highest average (1960) and largest minimum (1204). The infrared channels (3 & 4) also performed well, yielding an average and minimum divergence reduced by 0.3% and 2.8% respectively. Poorest performance was displayed by the visible channels (1 & 2), producing an average value of 422 and a minimum of 18.

Even though final classification employed all four ERTS bands, clearly the IR bands alone would produce essentially the same classification. The redundant information content of channels 1 and 2 suggests that spectral classification schemes based wholly on these bands would be unsuccessful.

In the last step of the analysis sequence, the classification algorithm, in conjunction with the printresults function, generated a map-like display (Figure 12). Each spectral class was depicted by a user-defined alphanumeric symbol. Overall training field classification performance was 96.0%; a figure which indicates the pattern recognition algorithm encountered very little confusion in mapping training field data. Application of an arbitrary threshold value (2.0) easily differentiated materials dissimilar to the mapped forest classes. Lake Drummond, in the swamp's interior, was correctly thresholded, as were U.S. route 460 north of the swamp and agricultural fields located along its eastern, western and southern boundaries.

Table 3. Summary of feature selection output for unsupervised classification (Ranked according to minimum divergence (DIJ(MIN))).

| CHANNELS | | | | DIJ(MIN) | D(AVE) |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 1204 | 1960 |
| - | 2 | 3 | 4 | 1189 | 1958 |
| 1 | - | 3 | 4 | 1172 | 1956 |
| - | - | 3 | 4 | 1169 | 1954 |
| 1 | 2 | - | 4 | 178 | 1661 |
| 1 | 2 | 3 | - | 161 | 1778 |
| 1 | - | - | 4 | 129 | 1625 |
| - | 2 | 3 | - | 112 | 1756 |
| - | 2 | - | 4 | 94 | 1630 |
| 1 | - | 3 | - | 71 | 1751 |
| 1 | 2 | - | - | 18 | 432 |

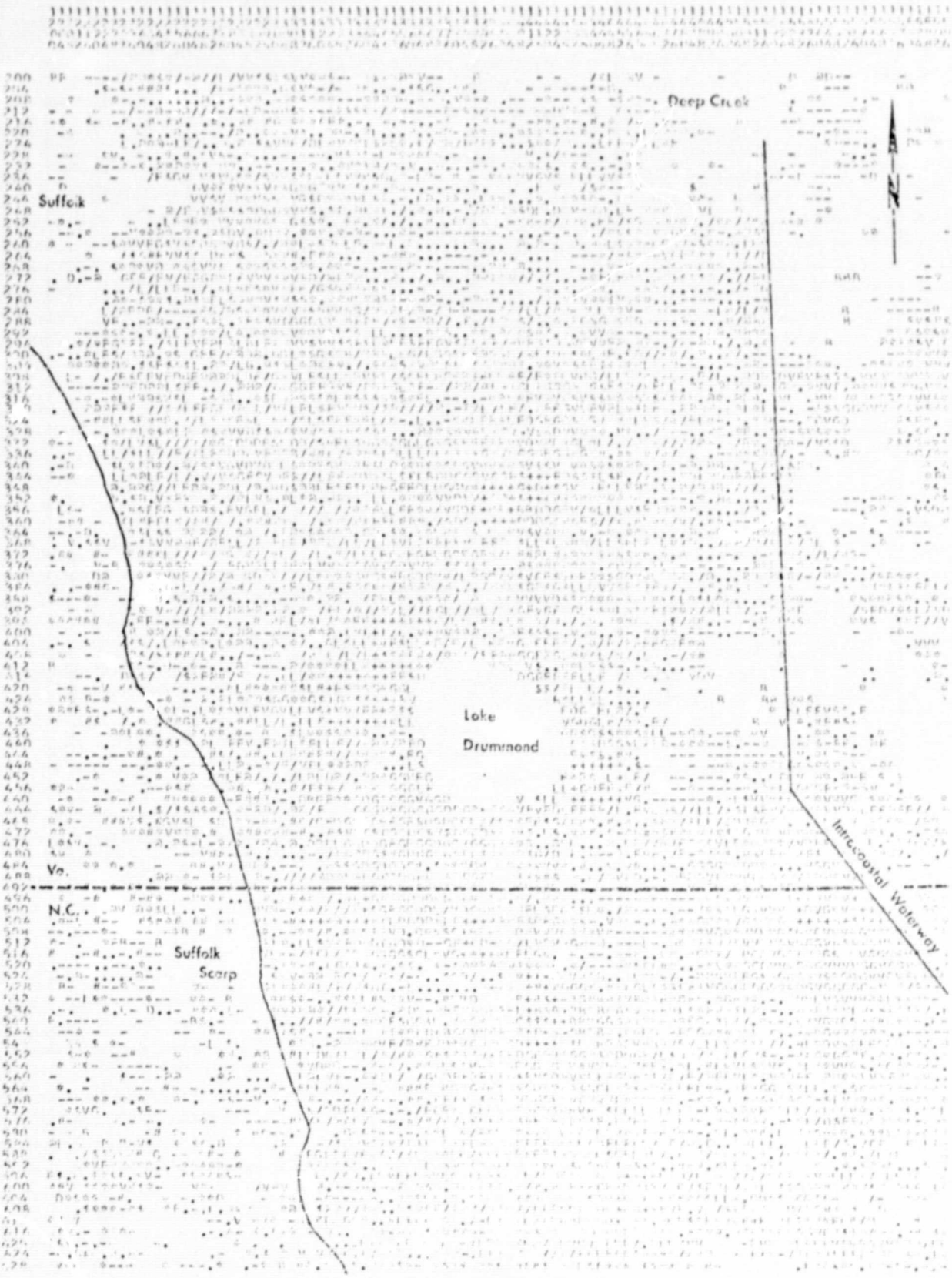



Figure 12. Unsupervised classification map of the Dismal Swamp. Each map symbol corresponds to a specific spectral class. (Thresholded pixels occur as blank areas).

Subsequent classification map analysis centered upon determination of the correspondence between spectrally similar ground cover (the map symbols) and categories of informational value, e.g. deciduous or coniferous forest, stand density, etc. In an attempt to quantify classification results, test fields identical to those used in the supervised classification were analyzed. These fields represented three categories: 1) Atlantic white cedar (Chamaecyparis thyoides), 2) Loblolly pine (Pinus taeda), and 3) deciduous forest. Map symbols were counted within the test fields in order to determine which symbol groupings might best represent each category of interest. Symbol overlap between categories occurred; however, an average classification performance of 82.5% was achieved by grouping spectral classes 12 & 13 as Atlantic white cedar; classes 4, 9, and 11 as Loblolly pine, and classes 1-3, 5-8, and 10 as deciduous forest (Figure 13). Realistically, the performance figure given above is too high because deciduous and pine forest, which cover a large portion of the swamp, were classified with only 77.8% and 69.8% accuracy respectively. In contrast, though cedar's areal extent is small, its classification accuracy of 100.0% was given equal weight in the calculation of average performance; thus tending to produce an overestimation of classifier accuracy.

Of 89,275 data points classified within the swamp, (each point corresponds to approximately 1.1 acres) 5218 (5.8%) were classified as cedar; 17,841 (20.0%) were classified as pine; 63,203 (70.8%) were classified as deciduous forest; and 3013 (3.4%) were thresholded as not being sufficiently alike any one of the training classes to allow classification.

Dense, monospecific stands of Atlantic white cedar corresponded

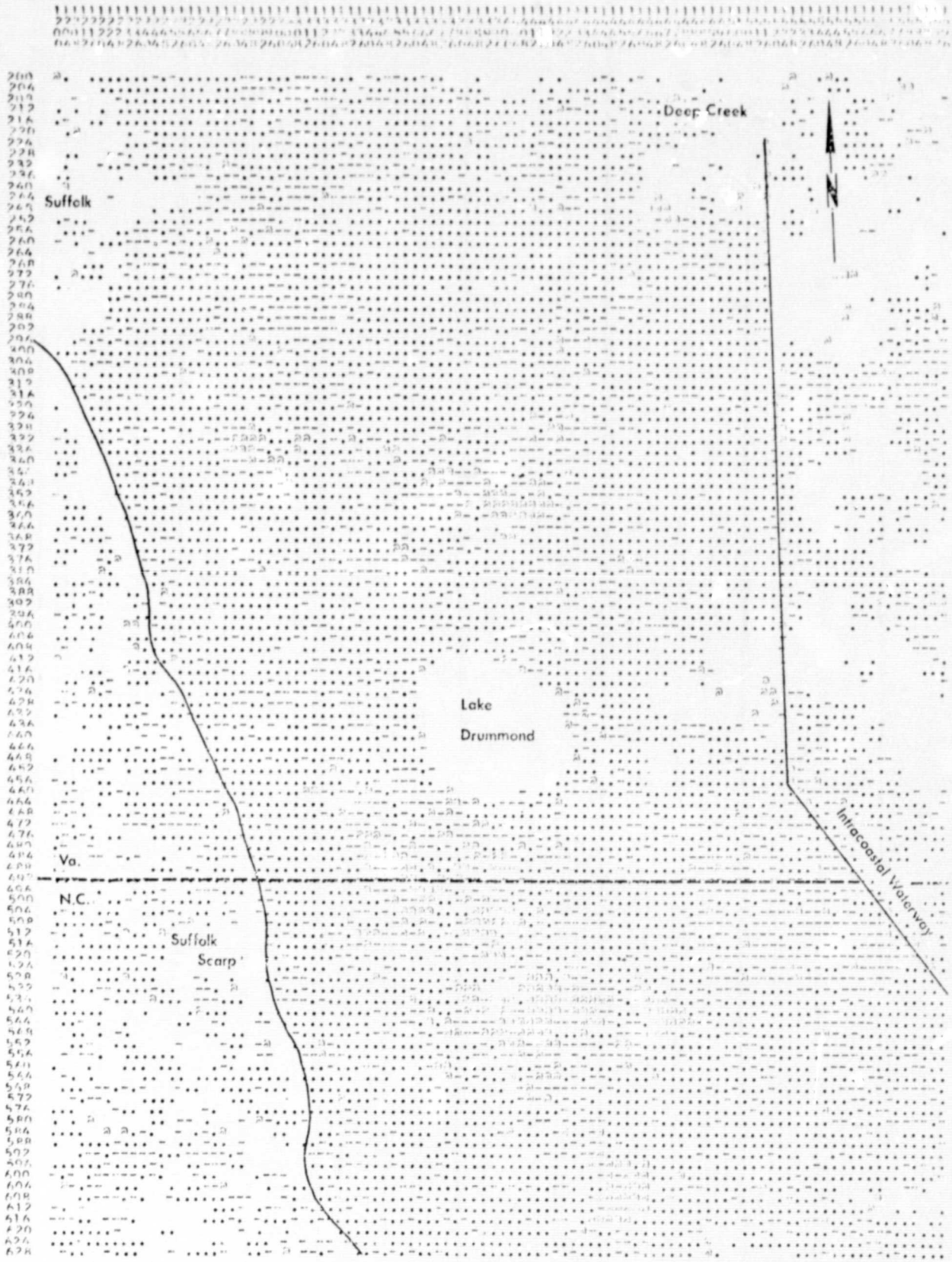



Figure 13. Unsupervised classification map produced by grouping spectral classes 12 & 13 as Atlantic white cedar; classes 4, 9, & 11 as Loblolly pine; and classes 1-3, 5-8, & 10 as deciduous forest. Cedar, pine, and deciduous forest are depicted as "@", "-", and "." respectively.

closely to spectral class 13, which exhibited the most absorbent qualities of the mapped spectral classes. Cedar misclassifications occurred, however, these were mainly confined to two areas. Several solitary incorrect mappings were noted around the shores of Lake Drummond, a phenomena which can be attributed to pixel averaging of water reflectance and forested terrain. Misclassifications also occurred within agricultural fields adjacent to the swamp. These incorrect mappings may have been due to the presence of standing water in the fields at the time of satellite data acquisition. Spectral class 12 most closely corresponded to forest stands dominated by Atlantic white cedar, but interspersed with a mixture of hardwoods; mostly Red maple (Acer rubrum) and Sweet gum (Liquidambar styraciflua). By mapping the combination of spectral classes 12 and 13, a good representation (100% test field accuracy) of the Atlantic white cedar present in the swamp was attained. Classification map interpretation suggested, however, that the extent of swamp cedar cover was slightly overestimated, a phenomena not reflected in the test field accuracy figure.

Test field analysis indicated that three spectral classes of intermediate reflectivity ( 4, 9, & 11 ) yielded the most accurate representation of Loblolly pine. As shown in Table 4, the classifier encountered confusion in separating pine from deciduous forest. This was because the three spectral classes chosen to represent pine overlapped with the deciduous forest category.

The swamp's deciduous forest was best represented by combining eight spectral classes of relatively high reflectivity (1-3, 5-8, & 10). This large number of classes is an indication of the comparatively wide range of reflectivities which correspond to deciduous cover. None of the swamp's plant communities associated with deciduous forest were uniquely mapped

Table 4. Unsupervised classification results, using all four channels.

CLASSIFICATION STUDY 515035761 CLASSIFIED MAY 30,1975

CLASSIFICATION TAPE/FILE NUMBER ... 225/ 1

CHANNELS USED

| | | | | | | |
|---|---|---|---|---|---|---|
| CHANNEL | 1 | SPECTRAL BAND | 0.50 TO | 0.60 MICROMETERS | CALIBRATION CODE = 1 | CO = 0.0 |
| CHANNEL | 2 | SPECTRAL BAND | 0.60 TO | 0.70 MICROMETERS | CALIBRATION CODE = 1 | CO = 0.0 |
| CHANNEL | 3 | SPECTRAL BAND | 0.70 TO | 0.80 MICROMETERS | CALIBRATION CODE = 1 | CO = 0.0 |
| CHANNEL | 4 | SPECTRAL BAND | 0.80 TO | 1.10 MICROMETERS | CALIBRATION CODE = 1 | CO = 0.0 |

TEST CLASS PERFORMANCE

| GROUP | NO OF SAMPS | PCT. CORCT | NUMBER OF SAMPLES CLASSIFIED INTO DECID | PINE | CEDAR | THRSHOLD |
|---|---|---|---|---|---|---|
| 1 DECID | 887 | 77.8 | 690 | 165 | 14 | 18 |
| 2 PINE | 235 | 69.8 | 58 | 164 | 11 | 2 |
| 3 CEDAR | 292 | 100.0 | 0 | 0 | 292 | 0 |
| TOTAL | 1414 | | 748 | 329 | 317 | 20 |

OVERALL PERFORMANCE( 1146/ 1414) = 81.0

AVERAGE PERFORMANCE BY CLASS( 247.6/ 3) = 82.5

by the classifier. Two deciduous forest classes (1 & 2) corresponded to the evergreen shrub bog community located in the swamp's southern portion. Inability of the classifier to discriminate between this shrub bog community and deciduous forest was due to their similar reflectances in the visible and near infrared wavelengths.

Supervised Classification

Initial vegetational distribution ground truth was provided by winter infrared imagery acquired at approximately 60,000 feet (1:120,000); ground verification was employed to substantiate image interpretation. Study site training and test fields were designated using this imagery as well as gray scale printouts of the swamp. Training and test fields were established in five categories of interest: 1) Atlantic white cedar 2) Loblolly pine 3) deciduous forest 4) agriculture and 5) water. In order to determine the number of Gaussian classes within each category, the cluster algorithm was employed. Using separability data provided by the cluster function, marginally separable classes were combined. Following this training class refinement, Atlantic white cedar and Loblolly pine were each represented by four classes; deciduous forest and water each by five classes; and agriculture by six classes. The number of data points for each refined spectral class is shown in Table 5. Lindenlaub (1973) reported that in theory a lower limit on the number of training data points for any class is n + 1 (n being the number of channels used by the classifier). He further suggested that 10 n be used as a practical lower limit, with 20 n to 100 n being optimal. Using the criteria suggested by Lindenlaub, all classes, except two within water and two within cedar, contained sufficient data points for a four channel classification (n = 4).

In the subsequent analysis step, the mean vector and covariance matrix

Table 5. The number of clusters in each category before and after refinement and the number of data points in each refined cluster.

| Category | Original Cluster No. | Refined Cluster No. | No. of Data Points |
|---|---|---|---|
| Cedar | 1 | I | 34 |
| | 2 | II | 34 |
| | 3 | III | 61 |
| | 4 | | |
| | 5 | IV | 86 |
| | 6 | | |
| Pine | 1 | I | 95 |
| | 2 | | |
| | 3 | II | 68 |
| | 4 | III | 85 |
| | 5 | | |
| | 6 | IV | 44 |
| Deciduous | 1 | I | 206 |
| | 2 | | |
| | 3 | II | 336 |
| | 4 | | |
| | 5 | III | 113 |
| | 6 | IV | 257 |
| | 7 | | |
| | 8 | V | 91 |
| Water | 1 | I | 27 |
| | 2 | II | 190 |
| | 3 | | |
| | 4 | III | 83 |
| | 5 | IV | 78 |
| | 6 | V | 22 |
| Agriculture | 1 | I | 258 |
| | 2 | | |
| | 3 | II | 112 |
| | 4 | III | 103 |
| | 5 | | |
| | 6 | IV | 50 |
| | 7 | V | 76 |
| | 8 | VI | 118 |

of each of the twenty four classes were calculated. In addition to the statistics information, the statistics processor was requested to graph histograms of each training class in all four channels. In every case an approximate Gaussian distribution was displayed. A coincident spectral plot, also produced by the statistics processor, is shown in Figure 14. As depicted by this plot, channels 1 and 2 displayed the ability to distinguish between forest and agriculture, but were unable to differentiate categories within forest, e.g. pine, cedar and deciduous forest. In contrast, channels 3 and 4 demonstrated the ability to distinguish pine, cedar and deciduous forest, but were unable to differentiate between the general categories of forest and agriculture.

Feature selection was employed to calculate the transformed divergence between all class pairs. Designated class pairs were differentially weighed in this calculation. For example, a weight of zero was assigned to all class pairs within cedar because a classification error within this category was immaterial to classification accuracy (pine, deciduous forest, water and agriculture were treated similarly). Table 6 is a summary of the minimum and average interclass divergences for all channel combinations of 4, 3, and 2. The combination of four channels produced both the highest average divergence (1981) and largest minimum value (1215). Poorest discrimination was displayed by channels 1 and 2, which produced average and minimum divergences reduced by 17.2% and 97.4% respectively. In every channel combination investigated, minimum divergence occurred when determining the statistical distance between forest classes. Seven of the eleven minimum values occurred when calculating the divergence between Loblolly pine and deciduous forest; thus indicating the similarity of their spectral signatures using August imagery.

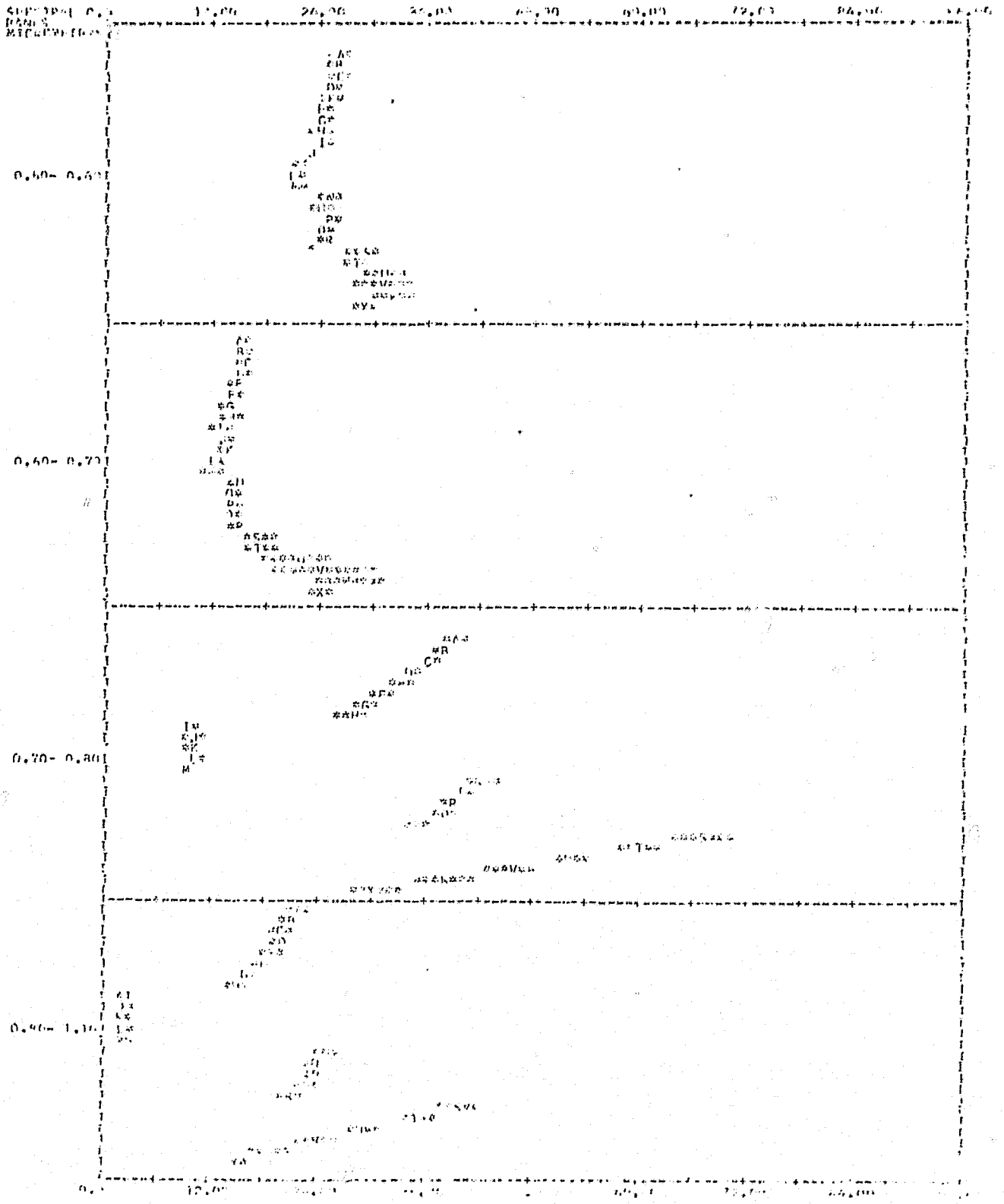


Figure 14. Coincident spectral plot. Spectral classes in pine, cedar, water, deciduous forest and agriculture are depicted as A-D, F-H, J-M, N-R, and T-X respectively.

Table 6. Summary of feature selection output for supervised classification (Ranked according to minimum divergence (DIJ(MIN))).

| Channels | | | | DIJ(MIN) | D(AVE) |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 1215 | 1981 |
| - | 2 | 3 | 4 | 995 | 1974 |
| 1 | 2 | 3 | - | 996 | 1963 |
| 1 | - | 3 | 4 | 958 | 1975 |
| 1 | 2 | - | 4 | 771 | 1952 |
| 1 | - | 3 | - | 726 | 1947 |
| - | - | 3 | 4 | 709 | 1951 |
| - | 2 | 3 | - | 502 | 1947 |
| 1 | - | - | 4 | 420 | 1937 |
| - | 2 | - | 4 | 373 | 1936 |
| 1 | 2 | - | - | 32 | 1640 |

Output of the classification function, combined with a printresults program, produced several classification maps of the Dismal Swamp (Figures 15-18). Alphanumeric symbols were chosen to depict the geographic location of various categories of interest, e.g. Atlantic white cedar, deciduous forest, coniferous forest, etc. These symbols were changed throughout the analysis in an attempt to bring about easier visual pattern recognition. Hand coloring of the printer output, though very time consuming, was also found effective as a means of increasing class pattern visibility. Of 89,275 data points classified within the swamp, 3746 (4.2%) were classified as cedar and 18,001 (20.2%) were classified as pine, i.e. 24.4% classed as coniferous forest. Deciduous forest was estimated to cover 69% of the swamp, almost three times the area of coniferous forest. This deciduous cover estimation did not differ significantly ($P = 0.05$) from an earlier estimation of 65% (Anonymous, 1974). Agricultural classifications accounted for 3.5% of the swamp's area. These classifications primarily corresponded to clear cut areas and natural forest openings, as well as recently burned areas. Only 3.2% of the data points were thresholded as not being sufficiently alike any one of the training classes to allow classification.

Five classifications, using selected channel combinations, were performed to determine the relative value of the visible (0.5-0.7 μm) and near IR wavelengths (0.7-1.1 μm) for mapping features of interest. Channel combinations utilized in this investigation were: 1) all four channels 2) channels 1, 2, 3 (one IR channel deleted) 3) channels 1, 3, 4 (one visible channel deleted) 4) channels 3, 4 (visible channels deleted) and 5) channels 1, 2 (near IR channels deleted). For each classification, an estimate of classifier accuracy was provided by test field performance.

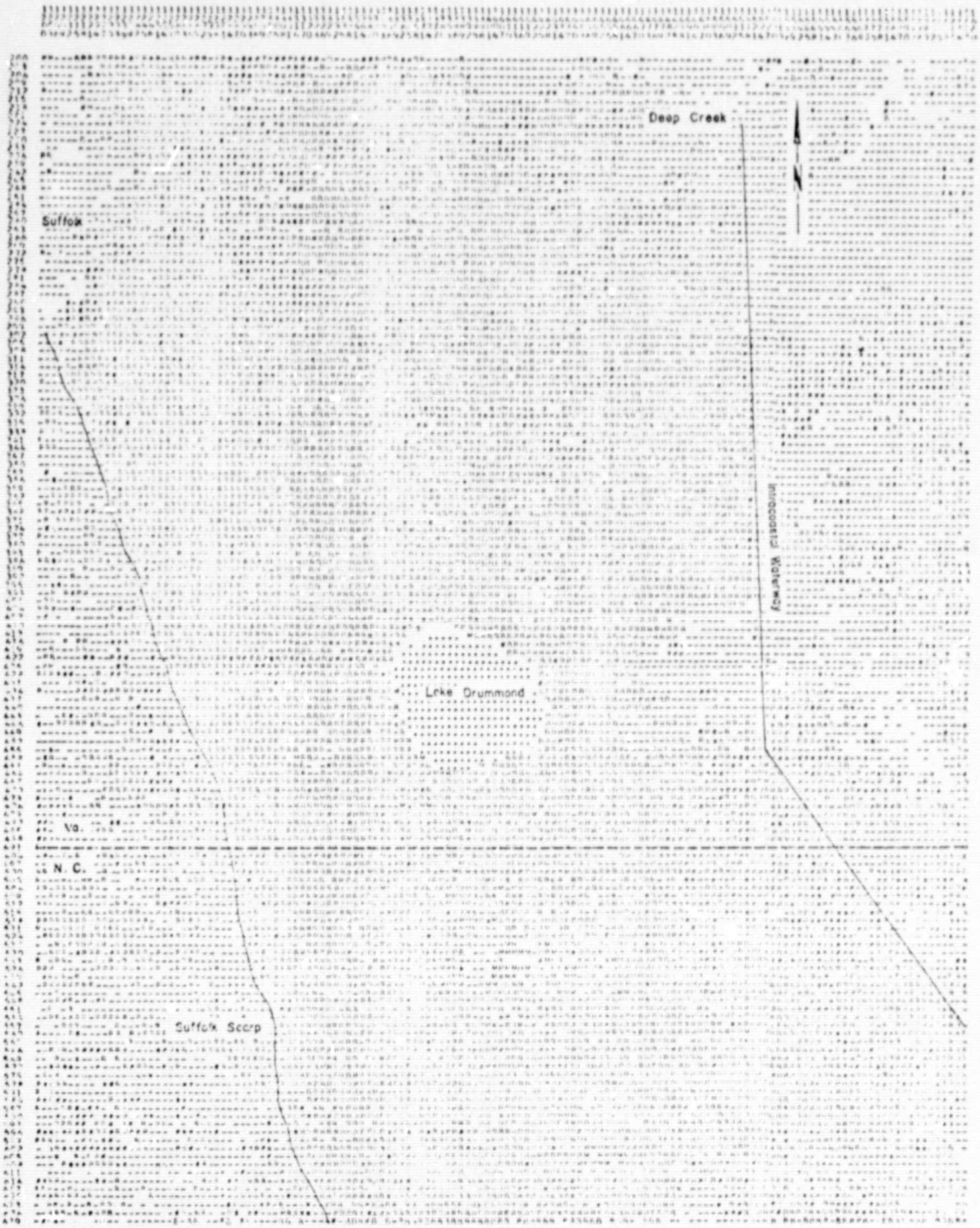



Figure 15. Supervised classification map of the Dismal Swamp. Categories are displayed as follows: cedar--"M", pine--"#", deciduous forest--"B", water--".", and agriculture--"-".


Figure 16. Supervised classification map of the Dismal Swamp. Atlantic white cedar's distribution is indicated by the symbol "X", all other categories appear as ".".

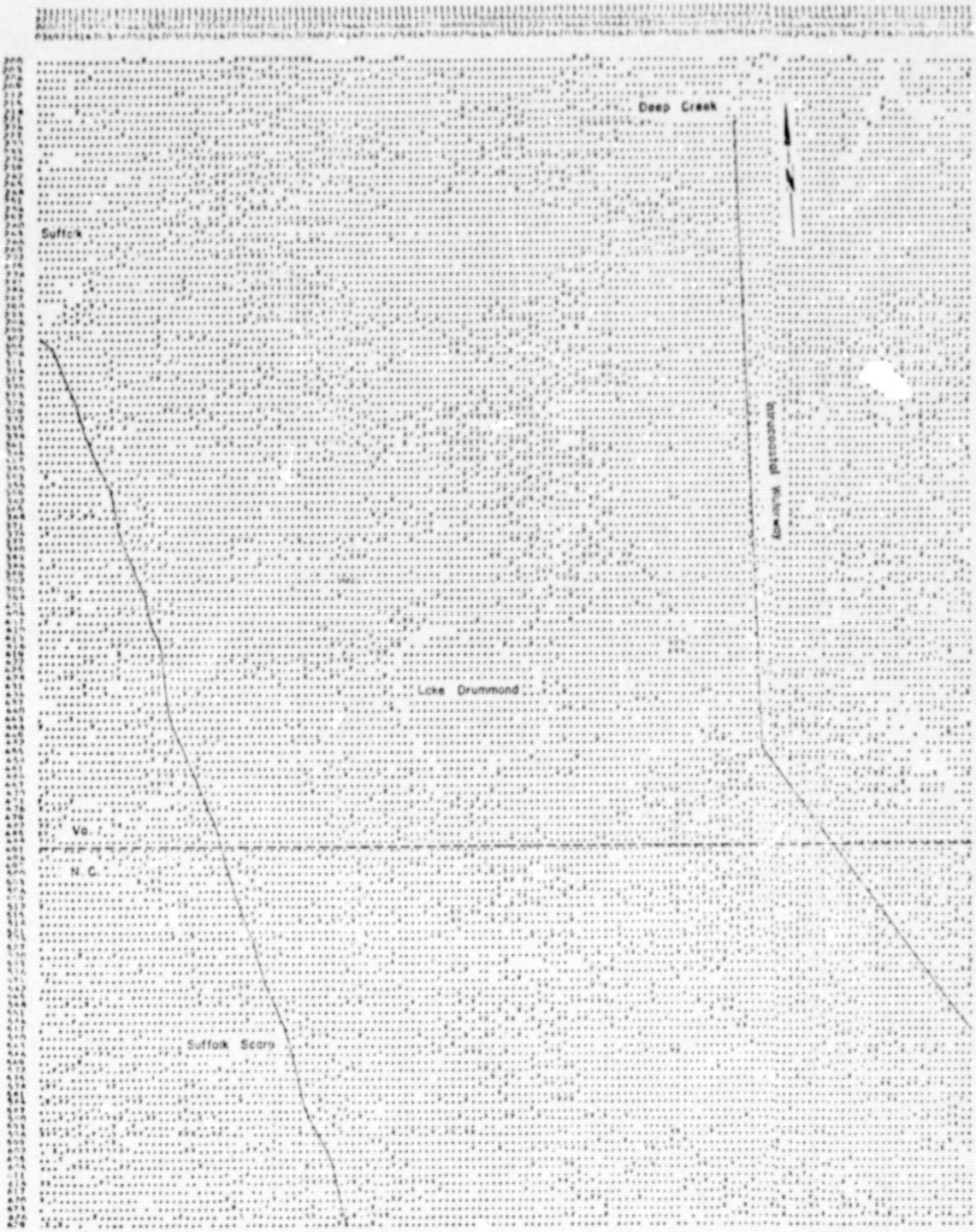



Figure 17. Supervised classification map of the Dismal Swamp. Loblolly pine's distribution is indicated by the symbol "X", all other categories appear as ".".

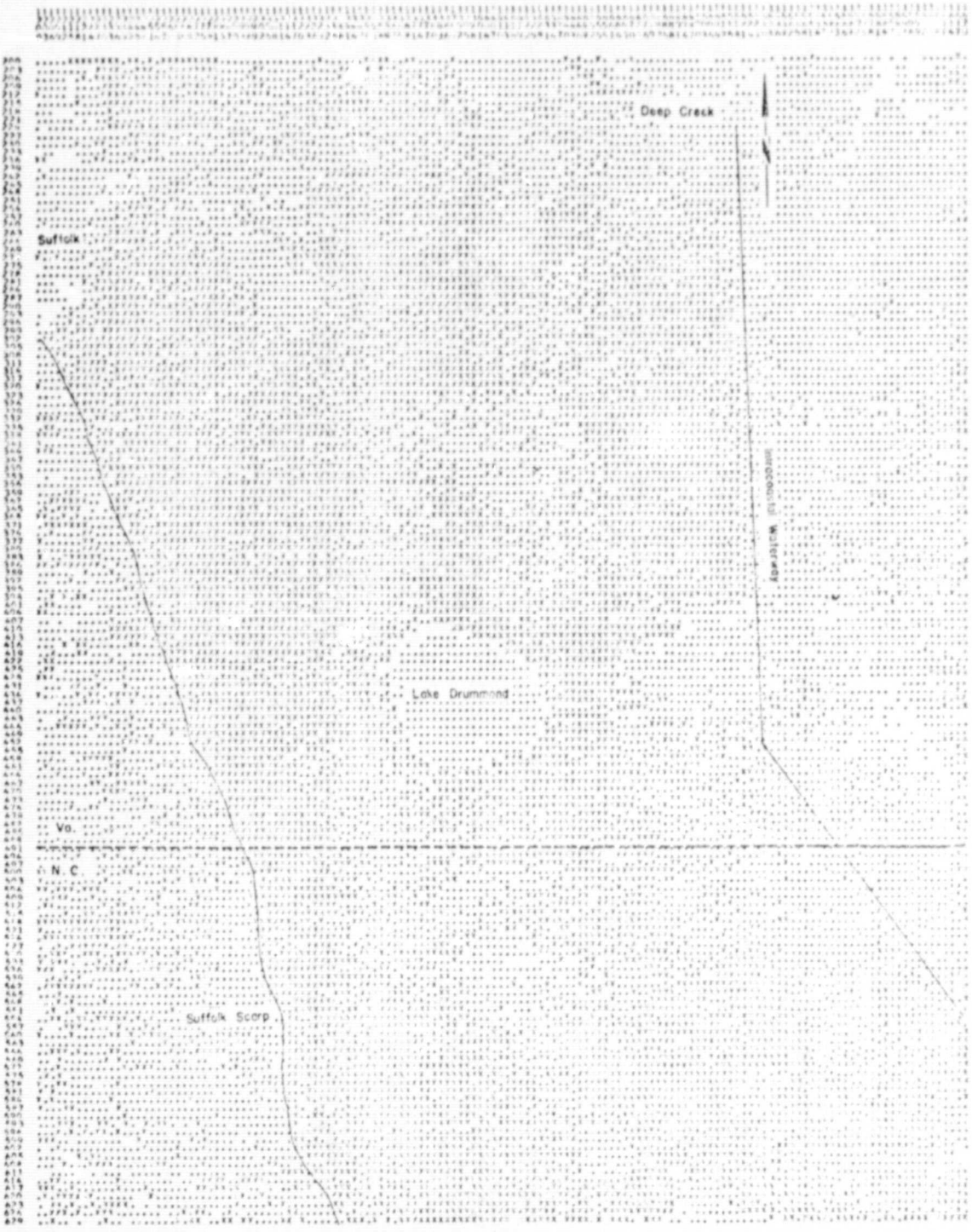



Figure 18. Supervised classification map of the Dismal Swamp. Deciduous forest's distribution is indicated by the symbol "X", all other categories appear as ".".

Tables 7-11 display the test class performance values obtained using each channel combination. From these data, Figures 19-26 were constructed which illustrate percent correct classification as a function of channels employed.

As indicated in Figure 19, an average test class accuracy of 88% (all categories considered) was equalled or exceeded by each of the three and four channel combinations. Deletion of the visible channels reduced classifier accuracy by approximately 6%. This reduction in accuracy was primarily due to confusion in the recognition of pine, deciduous forest and agriculture. The visible channels alone were much less accurate than the other channel combinations, producing an average accuracy value of 70.3%. This reduced accuracy was due primarily to inability of the classifier to distinguish between cedar and deciduous forest.

Evaluation of the ability of ERTS-1 data to accurately map combined forest (pine, cedar and deciduous forest) was a primary research objective. Figure 20 illustrates the degree to which the five channel combinations were able to distinguish combined forest from nonforest (water, agriculture). All channel combinations, except one, produced test accuracies in excess of 95%, indicating the relative ease with which this category was classified. Deletion of the visible channels reduced classifier accuracy to 90.5%, an accuracy loss due to the similar spectral responses of pine, cedar and agriculture within the near infrared wavelengths.

Figure 21 illustrates the combined conifer classification accuracies produced by the five channel combinations. Channels 1, 3 & 4 exhibited the highest classification accuracy (91.7%); however, channels 1, 2, 3 & 4 and channels 1, 2, & 3 performed within 5% of the best channel combination. Deletion of the visible channels in the classification caused a 10% reduction in test accuracy, brought about primarily by confusion in the classification

Table 7. Supervised classification results: all four channels considered.

CLASSIFICATION STUDY 51555A5283 CLASSIFIED ... NOV 15,1975

CLASSIFICATION PROFILE FILE ... 2221 1

CHANNELS USED

CHANNEL 1 SPECTRAL BAND 0.50 TO 0.60 MICROMETERS CALIBRATION CODE = 1 CC = 0.0

CHANNEL 2 SPECTRAL BAND 0.60 TO 0.70 MICROMETERS CALIBRATION CODE = 1 CC = 0.0

CHANNEL 3 SPECTRAL BAND 0.70 TO 0.80 MICROMETERS CALIBRATION CODE = 1 CC = 0.0

CHANNEL 4 SPECTRAL BAND 0.80 TO 1.10 MICROMETERS CALIBRATION CODE = 1 CC = 0.0

TEST CLASS PERFORMANCE

| GROUP | NO OF SAMPS | PCT. CORCT | NUMBER OF SAMPLES CLASSIFIED INTO: PINE | CEDAR | WATER | DECID | AGRIC | THRESHOLD |
|---|---|---|---|---|---|---|---|---|
| 1 BLUE | 235 | 82.6 | 194 | 2 | 0 | 33 | 0 | 4 |
| 2 CEDAR | 242 | 88.7 | 12 | 250 | 0 | 0 | 1 | 20 |
| 3 WATER | 442 | 97.7 | 0 | 0 | 432 | 0 | 0 | 10 |
| 4 DECID | 867 | 81.4 | 136 | 9 | 0 | 722 | 0 | 14 |
| 5 AGRIC | 1053 | 94.0 | 0 | 0 | 0 | 0 | 989 | 63 |
| TOTAL | 2989 | | 342 | 271 | 432 | 755 | 989 | 111 |

OVERALL PERFORMANCE/ 2597/ 2989) = 89.3

AVERAGE PERFORMANCE BY CLASS( 444.4/ 51 = 89.9

Table 8. Supervised classification results: one IR channel deleted.

Table 9. Supervised classification results: one visible channel deleted.

CLASSIFICATION STORY 5141326.15 CLASSIFIED MAY 21,1975

CLASSIFICATION TAPE/FILE NUMBER ... 2231 1

CLASSES USED

CHANNEL 1 SPECTRAL BAND 0.50 TO 0.60 MICROMETERS CALIBRATION CODE = 1 CO = 0.0

CHANNEL 2 SPECTRAL BAND 0.70 TO 0.80 MICROMETERS CALIBRATION CODE = 1 CO = 0.0

CHANNEL 4 SPECTRAL BAND 0.80 TO 1.10 MICROMETERS CALIBRATION CODE = 1 CO = 0.0

TEST CLASS PERFORMANCE

| CLASS | NO OF SAMPS | PCT. CORR. | NUMBER OF SAMPLES CLASSIFIED INTO: PINE | CEDAR | WATER | DECID | AGRIC | THRESHOLD |
|---|---|---|---|---|---|---|---|---|
| 1 PINE | 235 | 93.6 | 196 | 0 | 0 | 38 | 0 | 1 |
| 2 CEDAR | 292 | 93.5 | 16 | 273 | 0 | 0 | 3 | 2 |
| 3 WATER | 442 | 97.5 | 0 | 0 | 431 | 0 | 0 | 11 |
| 4 DECID | 887 | 78.9 | 145 | 7 | 0 | 706 | 3 | 12 |
| 5 AGRIC | 1053 | 93.3 | 15 | 2 | 0 | 0 | 982 | 54 |
| TOTAL | 2909 | | 372 | 282 | 431 | 744 | 988 | 80 |

OVERALL PERFORMANCE: 2582/ 2909) = 88.8

AVERAGE PERFORMANCE BY CLASS: 445.6/ 5) = 89.1

Table 10. Supervised classification results: only IR channels considered.

CLASSIFICATION STUDY [illegible] CLASSIFIED [illegible] MAY 31,1975

CLASSIFICATION [illegible] ... 2227 9

CHANNELS USED

CHANNEL 3 SPECTRAL BAND 0.70 TO 0.80 MICROMETERS CALIBRATION CODE = 1 GR = 0.0

CHANNEL 4 SPECTRAL BAND 0.80 TO 1.10 MICROMETERS CALIBRATION CODE = 1 GR = 0.0

TEST CLASS PERFORMANCE

NUMBER OF SAMPLES CLASSIFIED INTO

| SYMBOL | NUMBER OF SAMPLES | PCT. CORRECT | WATER | CEDAR | MIXED | DECID | AGRIC | THRESHOLD |
|---|---|---|---|---|---|---|---|---|
| 1 WATER | 255 | 63.9 | 163 | 0 | 0 | 68 | 6 | 0 |
| 2 CEDAR | 313 | 80.8 | 16 | 253 | 0 | 0 | 24 | 0 |
| 3 MIXED | 447 | 98.5 | 0 | 1 | 440 | 0 | 3 | 2 |
| 4 DECID | 537 | 79.0 | 102 | 5 | 0 | 706 | 12 | 2 |
| 5 AGRIC | 1063 | 77.4 | 71 | 111 | 0 | 14 | 815 | 42 |
| TOTAL | 2615 | | 411 | 381 | 440 | 788 | 860 | 46 |

OVERALL WEIGHTED AVERAGE 2275/ 2615) = 87.0

AVERAGE PERFORMANCE BY CLASS( 411.87/ 5) = 82.4

Table 11. Supervised classification results: only visible channels considered.

CLASSIFICATION STUDY [illegible]

CLASSIFICATION TAPE/FILE NUMBER ... [illegible]

CHANNELS USED

CHANNEL 1 SPECTRAL BAND [illegible] CALIBRATION CODE = 1

CHANNEL 2 SPECTRAL BAND [illegible] CALIBRATION CODE = 1

TEST CLASS DEPENDENCE

| GROUP | NO. OF SAMPS | PCT. CORRECT | NUMBER OF SAMPLES CLASSIFIED INTO: DIRT | CEDAR | WATER | HARDWD | AGRIC | TWOSPECIES |
|---|---|---|---|---|---|---|---|---|
| 1 DIRT | 225 | [illegible] | 208 | 16 | 0 | 12 | 4 | 0 |
| 2 CEDAR | 202 | [illegible] | 74 | 115 | 22 | 77 | 2 | 1 |
| 3 WATER | 442 | [illegible] | 0 | 14 | 419 | 1 | 0 | 4 |
| 4 HARDWD | 497 | [illegible] | 214 | 274 | 37 | 202 | 4 | 1 |
| 5 AGRIC | 1043 | [illegible] | 16 | 0 | 0 | 0 | 1020 | 1 |
| TOTAL | 2409 | | 512 | 419 | 478 | 292 | 1040 | 17 |

OVERALL DEPENDENCE: [illegible] = 70.5

AVERAGE DEPENDENCE BY CLASS: [illegible] = 70.2

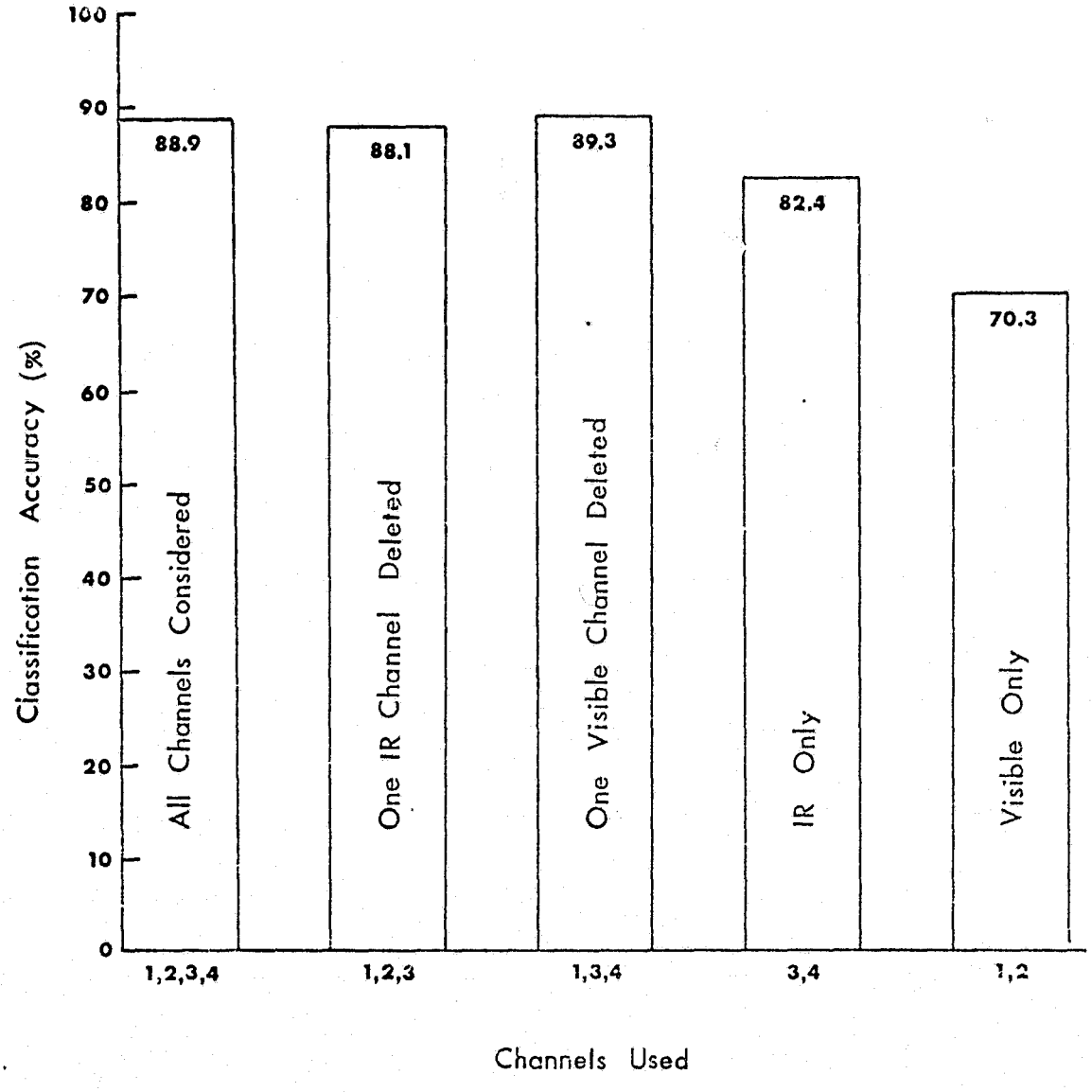



Figure 19. Influence of channel selection on classification accuracy of all categories, i.e. cedar, pine, deciduous forest, water and agriculture.

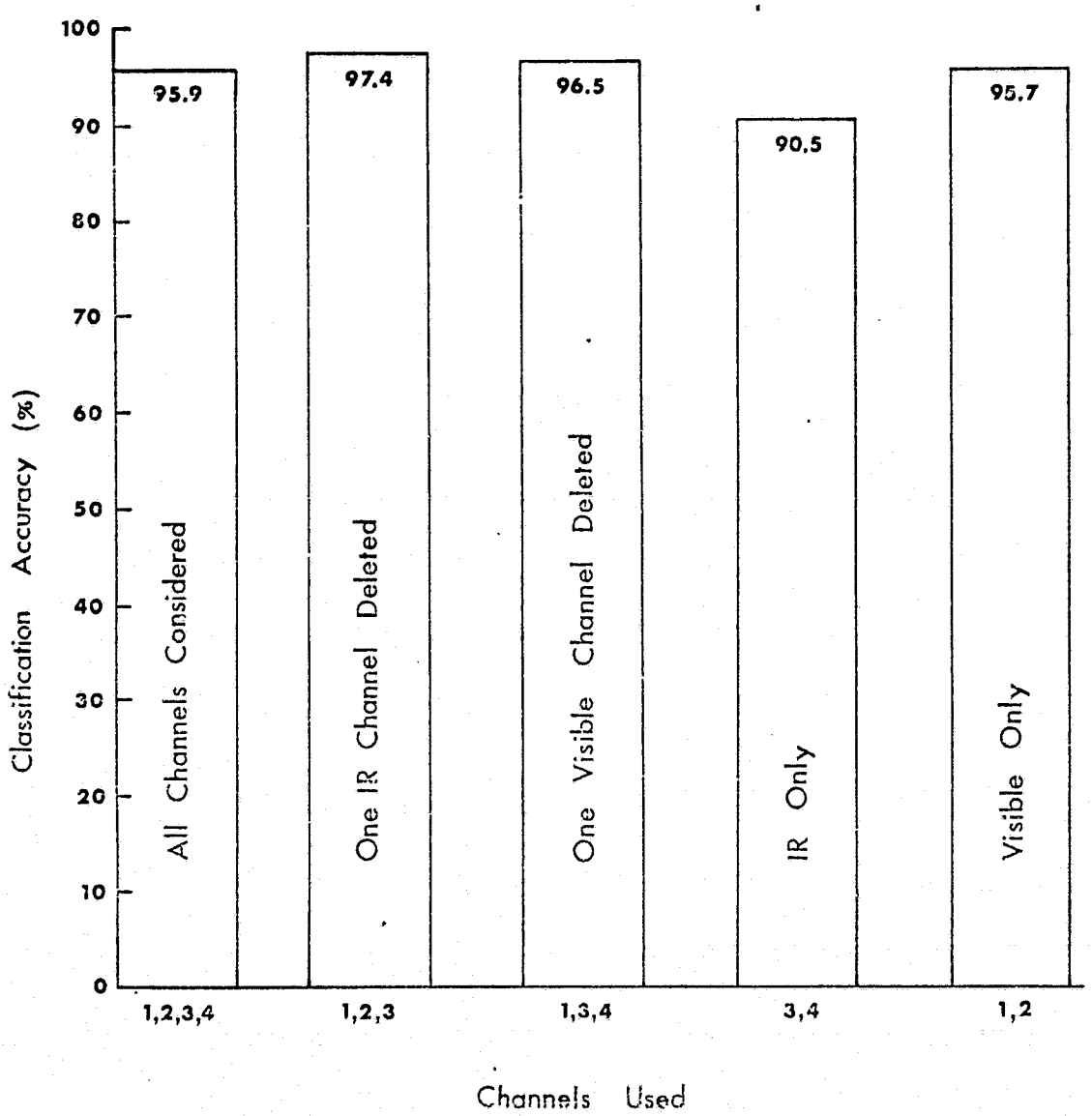



Figure 20. Influence of channel selection on combined forest classification accuracy.

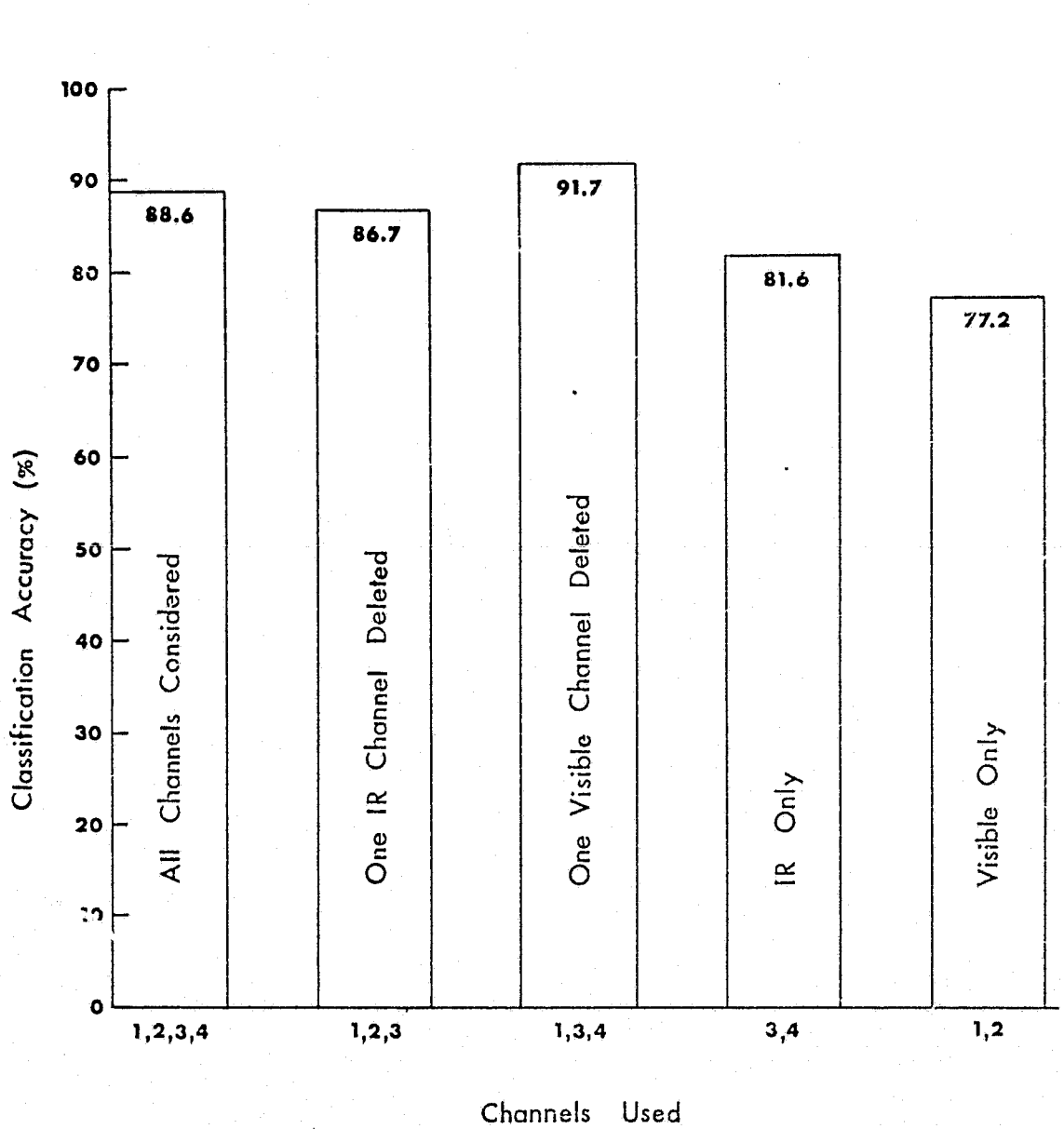



Figure 21. Influence of channel selection on conifer classification accuracy.

of conifers, deciduous forest and agriculture. The lowest test class performance was produced by the visible channels (77.2%), where decreased discrimination between conifers, deciduous forest and water accounted for the lowering of classification accuracy.

Loblolly pine was one of the most difficult of the forest categories to classify. As Figure 22 indicates, test performance accuracies exhibited by channels 1, 2, 3, & 4 and channels 1, 2, & 3 were within 5% of the highest accuracy (83.4%), produced by channels 1, 3 & 4. Classification performed using the infrared channels alone encountered confusion between the pine and deciduous forest categories, thereby reducing classification accuracy to 69.8%. Utilization of the visible channels alone produced an unrealistic test accuracy of 88.5%. This value is misleading because comparison of the classification map with color IR photography indicated the extent of pine within the swamp was grossly overestimated.

Atlantic white cedar test accuracies indicate this conifer was the most accurately classified of the forest categories. These high accuracies are the result of cedar stand homogeneity and the distinctive spectral signature which characterizes these stands. Figure 23 shows that channels 1, 3, & 4 exhibited the highest classification accuracy (93.5%), however, three other channel combinations, including the IR channel combination, produced similar test accuracies. Deletion of the IR channels reduced test accuracy to an extremely low level (38.7%).

Greatest classification difficulty occurred in mapping the swamp's deciduous forest. Figure 24 illustrates the influence of channel selection on classification accuracy. As indicated, four out of five channel combinations exhibited test accuracies in excess of 78%, with the highest accuracy (81.4%) being produced by all four channels. It is interesting to

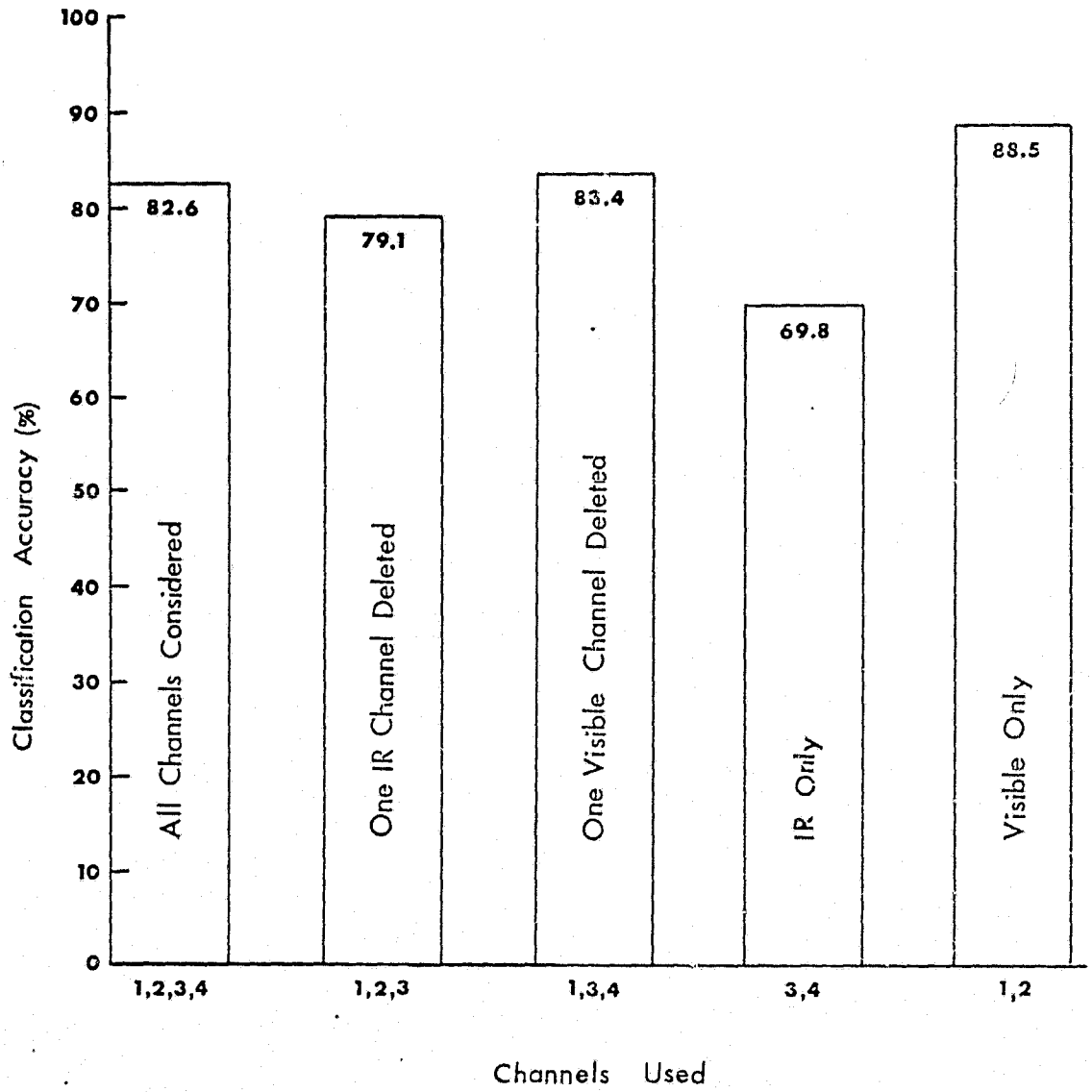



Figure 22. Influence of channel selection on Loblolly pine classification accuracy.

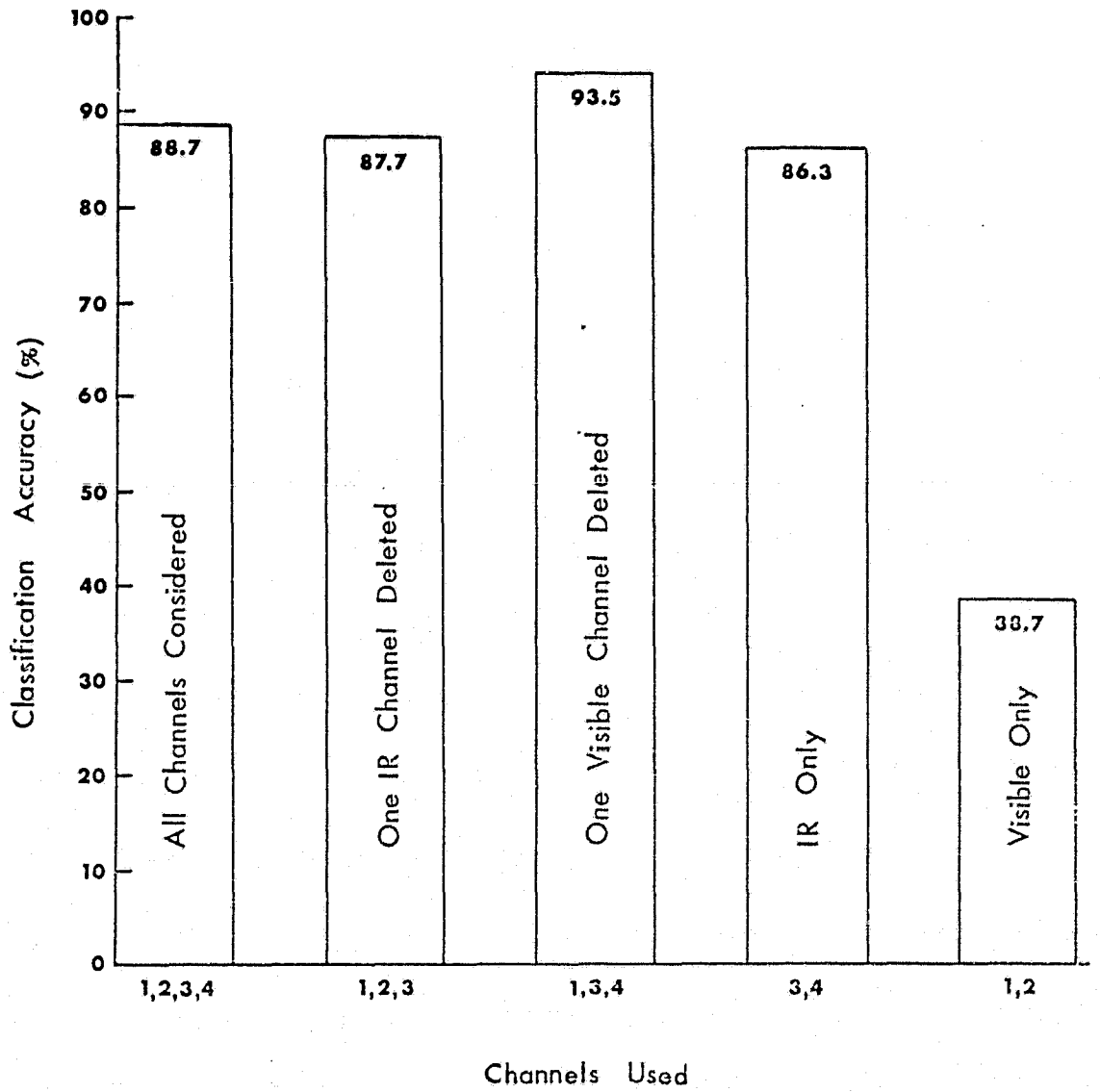



Figure 23. Influence of channel selection on Atlantic white cedar classification accuracy.

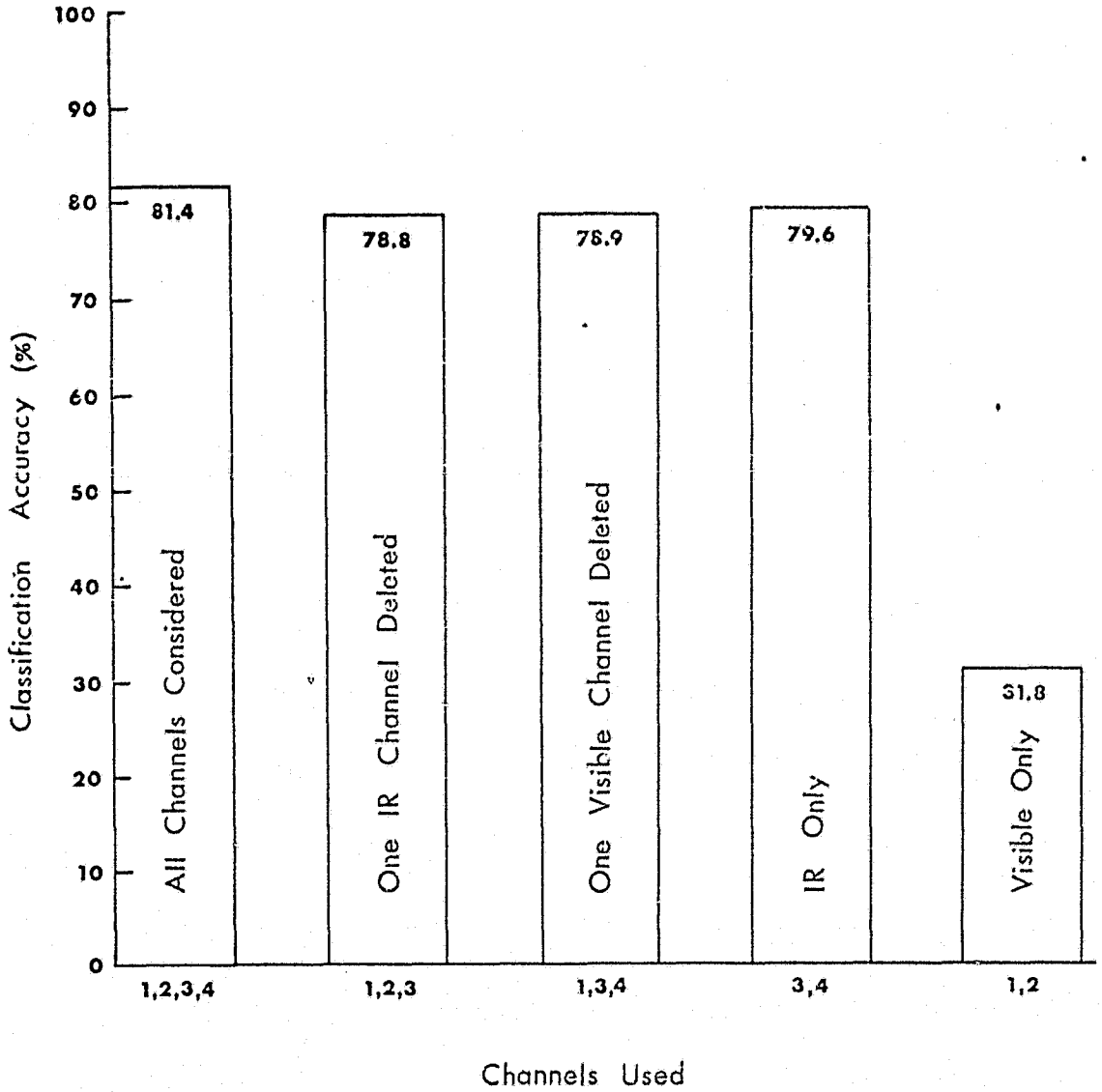



Figure 24. Influence of channel selection on deciduous forest classification accuracy.

note that deletion of the visible wavelengths did not degrade deciduous forest classification accuracy as it did with Loblolly pine. Loss of the infrared wavelengths, however, caused a precipitous drop in test accuracy to 31.8%, thus demonstrating the importance of the IR wavelengths in deciduous forest classifications.

Each of the five channel combinations classified water within the swamp (Lake Drummond) with a high degree of accuracy. As Figure 25 indicates, all channel combinations, except one, exhibited classification accuracies in excess of 97%. The two visible channels alone produced a slightly lower accuracy of 94.8%, brought about by confusion between cedar and water.

As Figure 26 illustrates, agricultural areas adjacent to the swamp were generally classified with a high degree of accuracy. Channels 1 and 2 attained the highest classification accuracy (97.7%) of the channel combinations investigated. Also performing well were the combinations of four and three channels, producing test accuracies within 5% of the best channel combination. The near infrared channels alone were least effective in classifying agriculture, producing a test accuracy of 77.4%. This accuracy loss was primarily due to inability of the classifier to discriminate between coniferous forest and agriculture.

Classification Problems

Problems encountered in mapping vegetation within the Dismal Swamp were often related to site complexity. The swamp's plant communities usually grade from one into another without sharply defined boundaries. This leads to a situation of natural randomness that makes classification into meaningful categories difficult. The only monospecific community that exists within the swamp is Atlantic white cedar, a fact clearly reflected in the high cedar classification accuracy. The other plant communities are

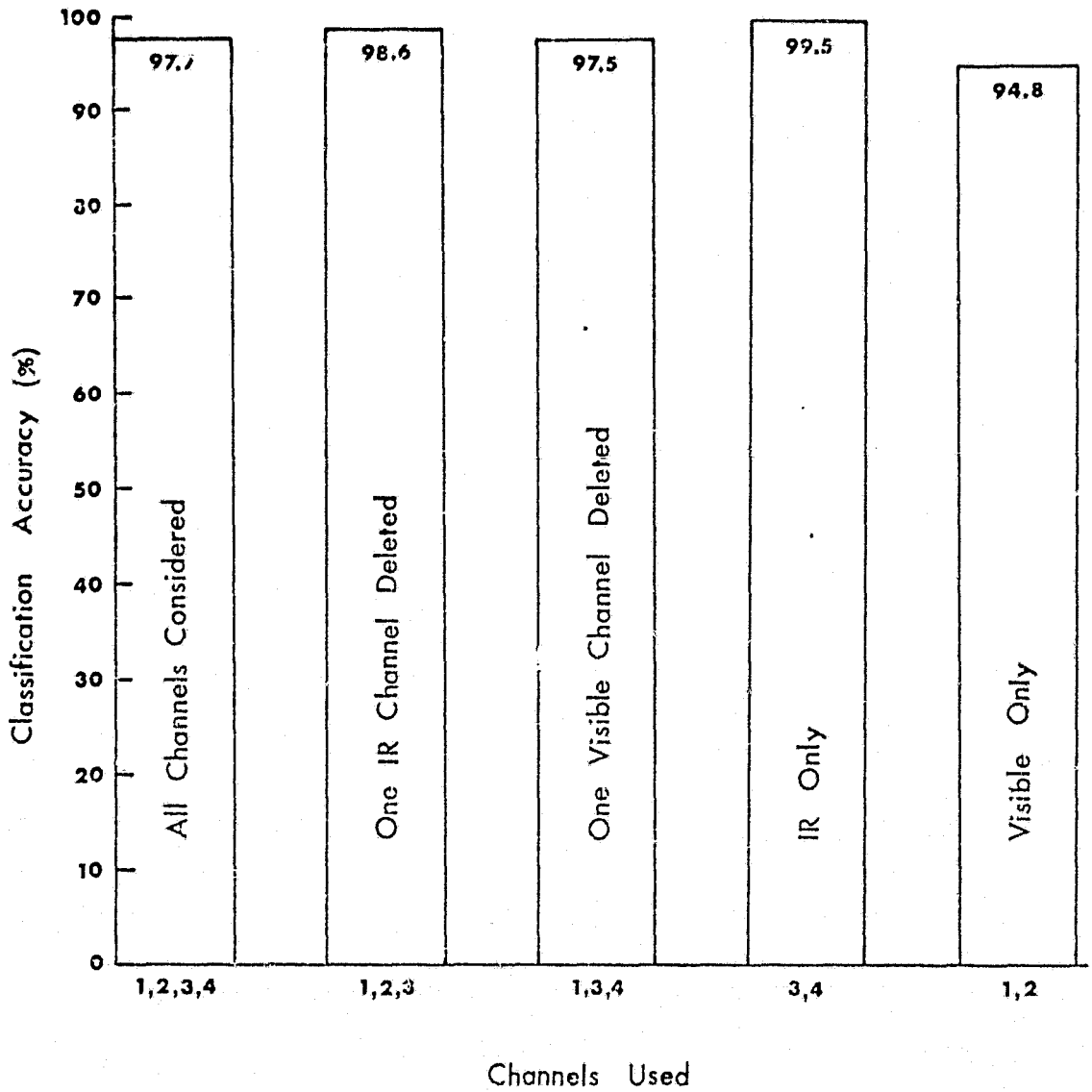



Figure 25. Influence of channel selection on water classification accuracy.

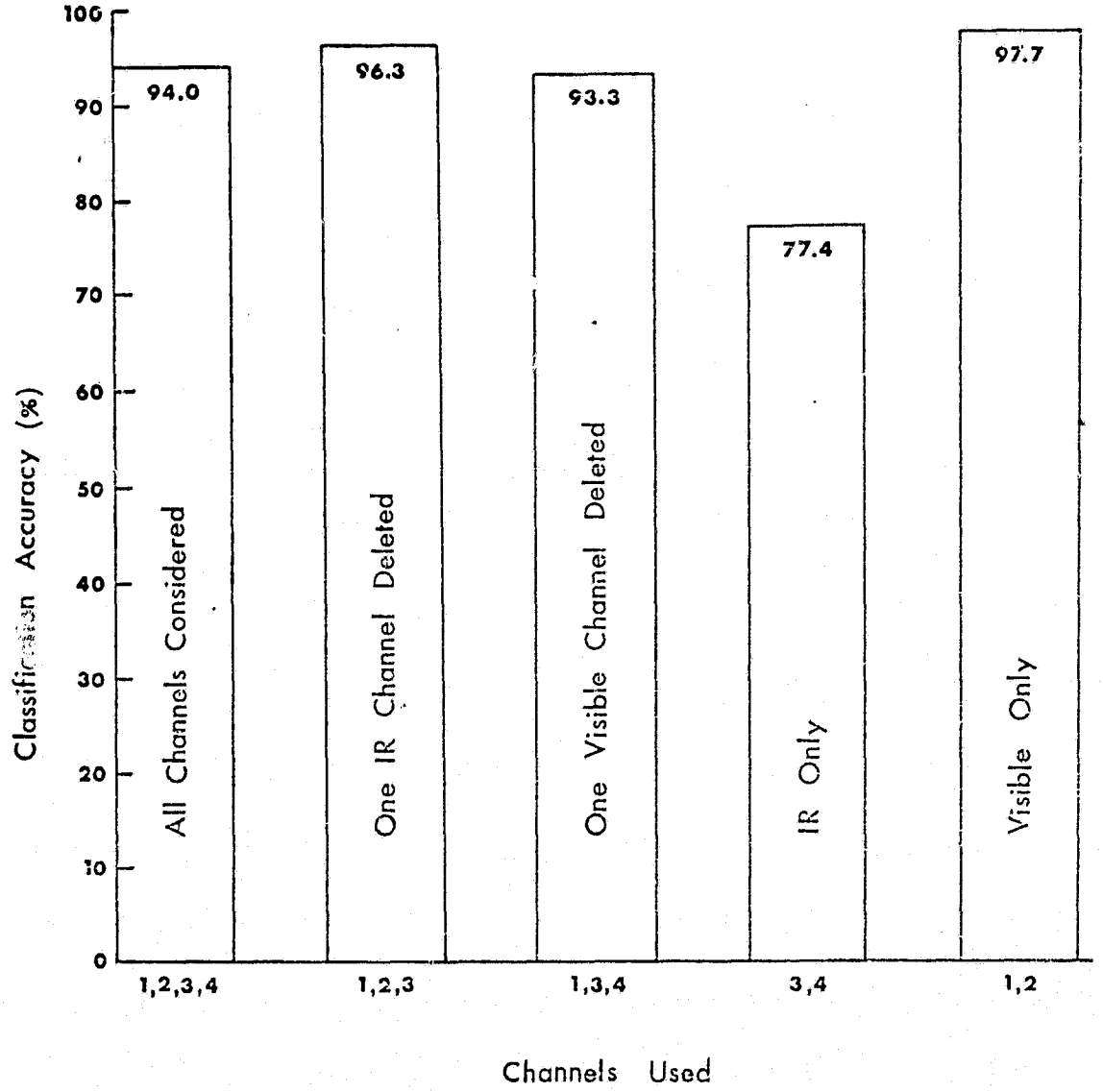



Figure 26. Influence of channel selection on agriculture classification accuracy.

characterized by the presence of unevenly distributed tree and/or shrub species. This greatly increases classification difficulty by adding to the problem of designating representative training and test areas.

Another difficulty inherent in mapping diverse vegetation with low resolution MSS data includes pixel averaging. Because tree and shrub species that comprise the swamp's plant communities are, of course, much smaller than the 80 meter square pixel size, and are distributed unevenly, spectral characteristics recorded by the scanner will not always be an accurate indicator of specific species composition. Hypothetically, at certain times of the year, a pixel containing a mixture of 25% Loblolly pine and 75% Red maple may have average reflectance characteristics identical to a pixel averaging the reflectances of 50% Yellow poplar (Liriodendron tulipifera), and 50% Sweet gum. Scanner data obtained during subsequent seasons of the year may remove this spectral signature ambiguity.

Problems were also encountered in attempting to locate, with any degree of certainty, the position of specific forested areas in ERTS-1 digital data. This was primarily due to: 1) the northeast-southwest skew contained in ERTS-1 MSS data 2) lack of large, identifiable, land features within the swamp (except Lake Drummond) and 3) rectangular computer output format. These problems were solved to a degree, in that locations in the ground scene can be generated by an affine mapping transformation developed by Blais (1975).

# CONCLUSIONS

## Swamp Vegetation Mapping Capability

ERTS-1 MSS data and LARS' automatic data processing techniques were applied to mapping swamp forest cover. Supervised classification accuracies in excess of 95% were attained which illustrate the classifier's forest cover mapping ability. The swamp's coniferous forests were classified well, producing an average test accuracy of 92%. Deciduous forest within the swamp was differentiated less successfully, yielding an 81% average test accuracy. The primary reason for this reduced classifier accuracy was confusion between deciduous forest and Loblolly pine. Classification performance could probably be improved by analyzing data acquired during the winter months, when differences in coniferous and deciduous reflectance spectra are maximized.

Atlantic white cedar, floristically the least complex plant community within the swamp, was very accurately mapped by the classification algorithm. Additionally, the swamp's evergreen shrub bog community was located using unsupervised classification techniques, though it was not mapped as a unique feature by the classifier. Discrimination of the swamp's deciduous plant communities was not possible due to the similarity of their summer spectral reflectances. Successful discrimination of these phytocommunities will probably require an indirect classification method, such as determination of water availability within the swamp during the winter months. Using this technique, the probable location of hydric and mesic deciduous plant communities could be determined.

## Channel Combination Study

Results indicate both the visible and near infrared wavelengths are valuable for mapping swamp vegetation. The near IR wavelengths were most

useful in differentiating forest categories, e.g. Loblolly pine, Atlantic white cedar and deciduous forest, but sometimes confused forest cover classes with other categories. The visible wavelengths alone successfully differentiated forest from nonforest, but performed very poorly when classifying pine, cedar and deciduous forest. Addition of one IR channel to the visible channels or one visible channel to the IR channels, substantially increased classification accuracy. Utilization of the spectral information in all four ERTS-1 channels did not appreciably increase classifier accuracy.

## LITERATURE CITED

Blais, R. N., G. E. Copeland, and T. Lerner. 1975. Use of LARS system for quantitative determination of smoke plume lateral diffusion coefficients from ERTS images of Virginia. Fourth Annual Remote Sensing of Earth Res. Conf., Univ. of Tenn. Space Inst.

Coggeshall, M. E. and R. M. Hoffer. 1973. Basic forest cover mapping using digitized remote sensor data and ADP techniques. LARS info. note 030573. The laboratory for applications of remote sensing, Purdue Univ. 131 pp.

Erb, R. B. 1973. The utility of ERTS-1 data for applications in agriculture and forestry. Proc. 3rd Earth Res. Tech. Satellite-1 Symp. U. S. Gov. Print. Ofc. NASA SP-327. 1:75-85.

Gates, D. M. 1965. Energy, plants, and ecology. Ecology 46(1&2):1-13.

Gates, D. M., H. J. Keegan, J. C. Schleter, and V. R. Weidner. 1965. Spectral properties of plants. Appl. Opt. 4(1):11-20.

Gates, D. M. 1967. Remote sensing for the biologist. Bioscience 17(5): 303-307.

Gates, D. M. 1970. Physical and physiological properties of plants. In Remote Sensing with Special Ref. to Agriculture and Forestry. Natl. Acad. Sci., J. R. Shay (ed.) Wash. D.C. pp. 224-251.

Gausman, H. W., W. A. Allen, and R. Cardenas. 1969. Reflectance of cotton leaves and their structure. Remote Sensing Environ. 1:19-22.

Heath, G. R. and H. D. Parker. 1973. Forest and range mapping in the Houston area with ERTS-1. In Proc. Symp. on Significant Results obtained from ERTS-1. U. S. Gov. Print. Ofc. NASA SP-327 1:167-172.

Hoffer, R. M. and C. J. Johannsen. 1969. Ecological principles in spectral signature analysis. In Remote Sensing in Ecology. Univ. Ga. Press. P. Johnson (ed.) Athens, Ga. pp. 1-16.

Hoffer, R. M., P. E. Anuta, and T. L. Phillips. 1972. ADP, multiband and multiemulsion digitized photos. Photogramm. Eng. 38(10): 989-1001.

Holter, M. R. 1970. Research needs: The influence of discrimination, data processing, and system design. In Remote Sensing with Special Ref. to Agriculture and Forestry. Natl. Acad. Sci., J. R. Shay (ed.) Wash. D.C. pp. 354-421.

Kirvida, L. and G. R. Johnson. 1973. Automatic interpretation of ERTS data for forest mgmt. Symp. on Significant Results obtained from ERTS-1. 1:1075-1082.

Knipling, E. B. 1969. Leaf reflectance and image formation on color infrared film. In Remote Sensing in Ecology. Univ. Ga. Press. P. Johnson (ed.) Athens, Ga. pp. 17-29.

Kodak. 1972. Applied infrared photography. Eastman Kodak Co.

Laboratory for Agricultural Remote Sensing (LARS). 1970. Remote multispectral sensing in agriculture. Purdue Univ. Agr. Exp. Sta. Res. Bull. 873. Vol. IV. 112 pp.

Lindenlaub, J. C. 1972. Remote sensing analysis: a basic preparation. LARS information note 110471. The laboratory for applications of remote sensing. Purdue Univ. 90 pp.

Lindenlaub, J. C. 1973. Guide to multispectral data analysis using LARSYS. LARS information note 062873. The laboratory for applications of remote sensing. 88 pp.

Meanley, B. 1973. The great Dismal Swamp. Audubon naturalist society of the central Atlantic states, Inc. 48 pp.

Myers, V. I., C. L. Wiegand, M. D. Heilman, and J. R. Thomas. 1966. Remote sensing in soil and water conservation research. Proc. 4th Symp. Remote Sensing Environ. Univ. of Mich., Ann Arbor, Mich. pp. 801-813.

Myers, V. I. 1970. Soil, water, and plant relations. In Remote Sensing with Special Reference to Agriculture and Forestry. Natl. Acad. Sci., J. R. Shay (ed.) Wash. D.C. pp. 253-297.

NASA. 1972. Earth resources technology satellite-data users handbook. Goddard Space Flight Center, G. E. document no. 71SD4249.

Stoner, E. R. and M. F. Baumgardner. 1972. Multispectral determination of vegetative cover in corn canopy. LARS print 111072. The laboratory for applications of remote sensing. Purdue Univ. 115 pp.

Swain, Philip H. 1973. Pattern Recognition: A Basis for Remote Sensing Data Analysis. LARS Information Note 111572, Purdue Univ.

Todd, W., P. Mausel, and M. F. Baumgardner. 1973. An analysis of Milwaukee county land use by machine-processing of ERTS data. LARS information note 022773. The laboratory for applications of remote sensing. Purdue Univ.

Woolley, J. T. 1971. Reflectance and transmittance of light by leaves. Plant physiol. 47:656-662.

Anonymous. 1974. Developmental history and ecology of the Dismal Swamp with recommendations for public ownership and management. Prepared by Bureau of Sports Fisheries and Wildlife, Northeast region and Dismal Swamp National Wildlife Refuge, Suffolk, Va.